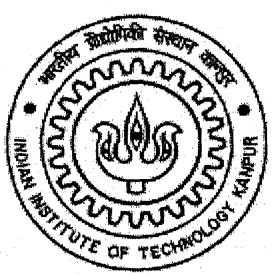
GROSS ERROR DETECTION IN STEADY STATE ONLINE MEASUREMENTS ON A CRUDE DISTILLATION UNIT

A Thesis submitted
in partial fulfilment of the requirements
for the Degree of

Master of Technology

by

Ankur Jain



to the

Department of Chemical Engineering
INDIAN INSTITUTE OF TECHNOLOGY, KANPUR

February 2001

CERTIFICATE

This is to certify that the work contained in the thesis entitled "**GROSS ERROR DETECTION IN STEADY STATE ONLINE MEASUREMENTS ON A CRUDE DISTILLATION UNIT**" has been carried out by ***Ankur Jain*** under my supervision and that it has not been submitted elsewhere for a degree.

February 2001

Dr. D. N. Saraf
Professor,
Dept. of Chemical Engineering,
Indian Institute of Technology,
Kanpur – 208 016, India

Dedicated

to

My Parents

Acknowledgements

To my thesis supervisor Dr. D. N. Saraf, for the valuable advice and encouragement that I got from him throughout. My gratitude to him is immense and cannot really be expressed in words. I consider it a rare privilege that I could work under him.

To all the professors of the department who showed the rigors and beauties of Chemical Engineering and encouraged me in all ways.

To all my lab mates. I was lucky to have lab mates like Palash, Tithi, Krishna Prasanth, Dhaval and Paritosh. The wonderful time I had here with my lab mates will be remembered, forever.

To my local guardian and Prince & Happy. I will never forget the nice times I spent with them. They never let me feel away from home. It was great to be a part of the home away from home.

To all my friends – within and outside IIT.

To IITK itself, for its stimulating academic environment. And also for gifting me with many special moments.

To my alma mater HBTI Kanpur. For those lovely memories that can sail one through all the rough rides in life.

To my sister, for being the great friend that she is to me.

Finally, to my group 'b in touch' which filled my mailbox with lots of fun.

Financial support from project C/CHT/CHE/95043 is gratefully acknowledged.

CONTENTS

		Page No.
List of Tables		**vii**
List of Figures		**viii**
Nomenclature		**ix**
Abstract		**xi**

Chapter

1	**Introduction**	**1**
	Crude distillation unit	3
	Present work	5
2	**Literature Review**	**6**
	Observations	8
3	**Generalized Likelihood Ratio Method For Gross Error Detection**	**10**
	Process Model	11
	The GLR Method	11
	Procedure for single gross error	11
	Strategy for multiple gross error	14
	Covariance matrix estimation	16
	Direct method	16
	Indirect method	16
	An example problem	18
4	**Mathematical Formulations**	**20**
	Gross error detection in flow rates	24
	Q matrix estimation	25
	Strategy for on-line gross error detection in flow rates	25

Gross error detection in temperatures 26

Assumptions 26

Q matrix estimation 29

Strategy for on-line gross error detection in temperatures 30

Strategy for on-line gross error detection in both, flow rates and temperatures 31

5 Results and Discussion 32

Single error in flow rate measurements 32

Two errors in flow rates measurements 34

Three errors in flow rates measurements 35

Single error in temperature measurements 36

Two errors in temperature measurements 37

Three errors in temperature measurements 38

Simultaneous errors in flow rate and temperature measurements 39

6 Conclusions and Recommendations 42

References 43

Appendix

A Data Sets for Q matrix Estimation 45

B A Brief Review of Graph Theory 47

Graph and Subgraph 47

Path cycle and connectivity 47

Trees, spanning trees, branches and chords 49

Cutsets and fundamental cutsets 49

Matrix representation of digraphs and graphs 51

Observability and redundancy 52

C Users Manual **53**

Q matrix generation 53

Input files 53

Output files 53

Procedure 53

Gross error detection in flow rates 54

Input files 54

Output files 54

Procedure 54

Gross error detection in temperatures 55

Input files 55

Output files 56

Procedure 56

D Program Listing **58**

List of Tables

Table 3.1	Results of GLR method applied to an example problem	19
Table 5.1	Single error in flow rate measurements	33
Table 5.2	Two errors in flow rate measurements	34
Table 5.3	Three errors in flow rate measurements	36
Table 5.4	Single error in temperature measurements	37
Table 5.5	Two errors in temperature measurements	38
Table 5.6	Three errors in temperature measurements	39
Table 5.7	One error in flow rate and one error in temperature measurements	40
Table 5.8	Two errors in flow rate and one error in temperature measurements	41
Table 5.9	One error in flow rate and two errors in temperature measurements	41

List of Figures

Figure 1.1	Steps involved in Data Reconciliation And Gross Error Detection	2
Figure 1.2	Crude Distillation Unit	4
Figure 3.1	Flow diagram for Ammonia Synthesis loop	18
Figure 4.1	Reduced Network for the Crude Distillation Unit	21
Figure A.1	A process network	48
Figure A.2	Graph G	48
Figure A.3	Subgraph of G	48
Figure A.4	A disconnected graph with two components	50
Figure A.5	A tree	50
Figure A.6	A spanning tree of G	50
Figure A.7	A Process Digraph	51
Figure A.8	Incidence matrix, M′	51

NOMENCLATURE

A	incidence matrix
A', A^T	transpose of A
b	magnitude of error
$\hat{b}$	estimated magnitude of b
C_{PL}	heat capacity of liquid
C_{PV}	heat capacity of vapour
e_i	vector with unity in position i and zero elsewhere
F	set of vectors Ae_i
$\mathbf{f}_i$	i^{th} element of set **F**
H	vapour enthalpy
H_0	null hypothesis
H_1	alternative hypothesis
ΔH_v	heat of vaporization
h	liquid enthalpy
$\mathbf{K_w}$	Watson characterization factor
M	molecular weight
m	no. of sets of measurements
Q	covariance matrix of measurements errors
$q_{i,j}$	elements of Q
R	universal gas constant
r	residual vector of measurements
Sg	specific gravity
T	GLR test statistics
T_c	critical temperature
V, H	covariance matrix of residuals
V^{-1}	inverse of V
v	vector of measurement errors
x	true values of state variables
z	measurement vector

Greek Symbols

μ	unknown expected value of r
λ	likelihood ratio
ε	prespecified threshold for GLR test
ω	acentric factor

Other Symbols

cov (.)	covariance of
E (.)	expected value of
Pr (.)	probability of
sup (.)	supremum of

Some Abbreviated Forms

UN	unstabilized naphtha
CR	circulation reflux
SCN	straight cut naphtha
SK	superior kerosene
LGO	light gas oil
HGO	heavy gas oil
SS	side stripper
LR	long residue
PA	pump around

ABSTRACT

Crude distillation is perhaps the most important unit operation carried out in a petroleum refinery. The raw data measured on-line, if used directly, may result in poor or erroneous control actions because of presence of random and/or gross errors. A software package, based on the Generalized Likelihood Ratio method, was developed to detect the gross errors in the flow rate and temperature measurements of a crude distillation unit (CDU) in steady state. Not all the variables involved in the problem are measured and in the absence of these data, the material and enthalpy balances could not be performed at the various nodes of the process diagraph. To overcome this difficulty a steady state process simulator using a non-linear analytical model of the CDU was coupled with the gross error detection algorithm.

The algorithm was tested offline using online data obtained from an operating refinery. The data were perturbed with known disturbances in one or more measurements. The algorithm was able to detect the presence of a single error in a flow rate as small as one percent of the stream flow. Presence of two or more errors rendered the algorithm less sensitive. Gross error detection in temperature measurements was less accurate since detection is dependent on enthalpy balance, which is an empirically correlated quantity. A single error of 3 to 5 percent in temperature measurement was easy to identify. The algorithm was able to detect upto three errors (in flow and/or temperature) measurements at a time reasonably well. However, as the number of errors increased, the performance of the algorithm was found to deteriorate.

Chapter 1

INTRODUCTION

Process data are the foundation of process monitoring, evaluation and control. Advancements in automation allow the collection of large volumes of process data. A chemical process may be equipped with scores or even hundreds of sensors with sampling intervals of seconds or minutes. A modern chemical plant consists of a large number of process units such as reaction vessels, distillation columns, storage tanks etc., which are interconnected together by a complicated network of streams. Measurements of mass flow rates, temperatures, and sometimes concentrations, and so forth are routinely made for the purpose of process control and process performance evaluation. These measurements are expected to satisfy mass and energy balance constraints associated with the process network when the process is in steady state. The given constraints are generally not satisfied, however because of presence of random errors and gross errors (called outliers) in the process data. While the former errors may arise because of random instrument noise, process noise, turbulence around the sensor etc, the later errors are due to miscalibrated or malfunctioning measuring instruments, unsuspected leaks, etc.

An additional difficulty arises because of the fact that not all variables are measured because of cost considerations or technical infeasibility. Therefore, it is necessary to adjust the measured variables and, if possible, estimate the unmeasured variables so that they satisfy the balance constraints; this is known as *data reconciliation problem.* Moreover, the adjustments in the process data should be utilized to detect the presence of gross error so that suitable corrective actions can be taken; which is called the *gross error detection problem.* If one or more gross error is present in the measurements, the adjustments made to all of the measurements will be strongly affected and, therefore, it must be detected and eliminated or corrected before valid data reconciliation can be performed and the statistical assumption of purely random error is not valid.

So a three step method should be adopted to solve this dual problem for a nonlinear steady state process. In the first step data reconciliation is performed assuming no gross error is present. Gross error detection is to be performed based on the linearized model

using one of the several available techniques. After eliminating all the measurements suspected of containing gross errors, final data reconciliation is performed to get the corrected estimates. Figure 1.1 shows the steps involved.

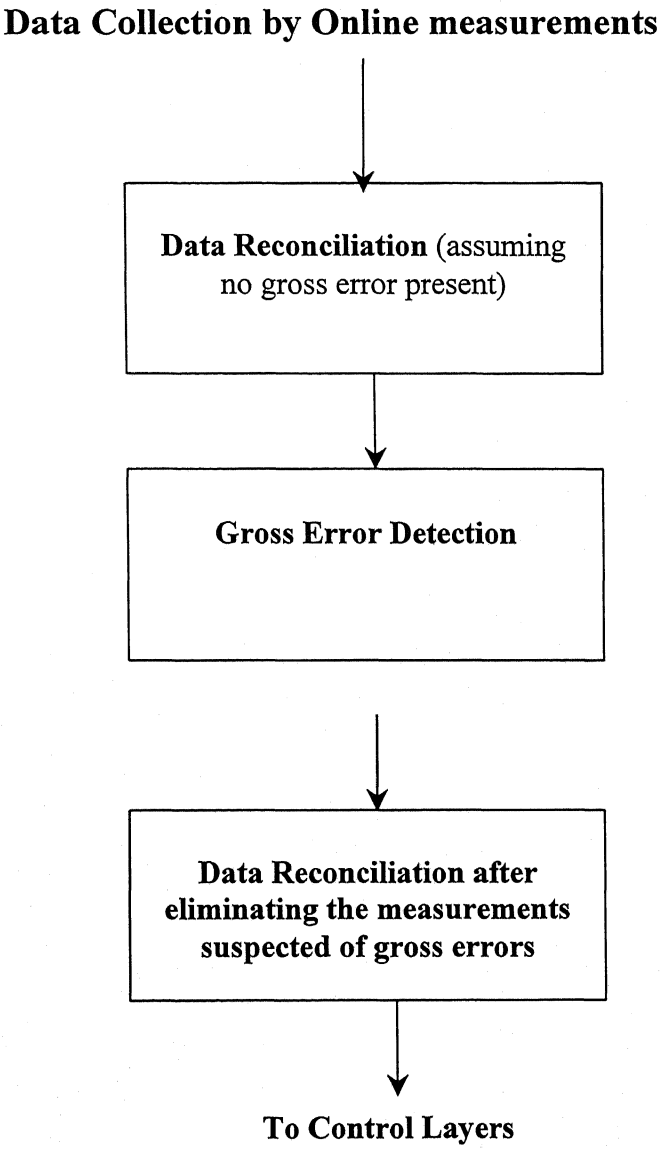


Figure 1.1: Steps involved in Data Reconciliation and Gross Error Detection

Crude distillation is a very important operation in any petroleum refinery. It is the process by which crude oil, which is a mixture of several hundred hydrocarbons of different molecular weights, is separated into different products according to their boiling ranges. Some of these products are blended into salable products whereas others

are treated in other units of the refinery subsequently. Most of the products of crude distillation are characterized by certain properties like Ried vapor pressure (RVP) for straight cut naphtha (SCN), flash point for naphtha, superior kerosene (SK)/ aviation turbine fuel (ATF), light gas oil (LGO), heavy gas oil (HGO) and long residue (LR), pour point for LGO, HGO. All these products need to conform to certain standards and this calls for prediction of these properties so that necessary control actions through feedback mechanism may be taken to prevent quality give away. Hence, there is a need for an estimation of the product properties online based on the measurements of the operating parameters, made in real time. In order that product properties can be reliably predicted the measured online data must be reconciled and gross error detected before these can be used for any application.

The methods of data reconciliation rely on redundancy of measurements (Mah and Tamhane, 1985) which is usually not available in process industry because of cost considerations. In absence of sufficient variables being measured, it is not possible to undertake data reconciliation for a crude distillation unit (Dutta, 2000). However, averaging of several measurements made in quick succession for each variable minimizes random error component. Hence, this study is directed at developing a suitable algorithm for gross error detection only.

Crude Distillation Unit (CDU)

A crude distillation unit in an operating refinery was used to obtain design and operating data for the present study (Figure 1.2). The CDU in question is 67.35 m high consisting of 49 trays and having three different sections with different diameters. It operates at about 2 to 3 times the atmospheric pressure. There are four side-strippers associated with the CDU column from which products – SCN, SK/ATF, LGO and HGO are drawn in succession. The numbers of trays in these four side-strippers are 6, 6, 6 and 4 respectively. There are four pumparounds (circulating refluxes), which, depending upon their position in the column, are referred to as TOP CR, SK CR, LGO CR and HGO CR. The bottom product is long residue (LR) and is fed to vacuum distillation unit (VDU) for further processing. The feed to CDU enters at 44^{th} tray. The overhead vapor from the top tray of the column enters a partial condenser. After condensation, the condensed stream goes to a reflux drum. A part of the liquid is further sub-cooled and is returned to the top plate as column reflux, while the rest,

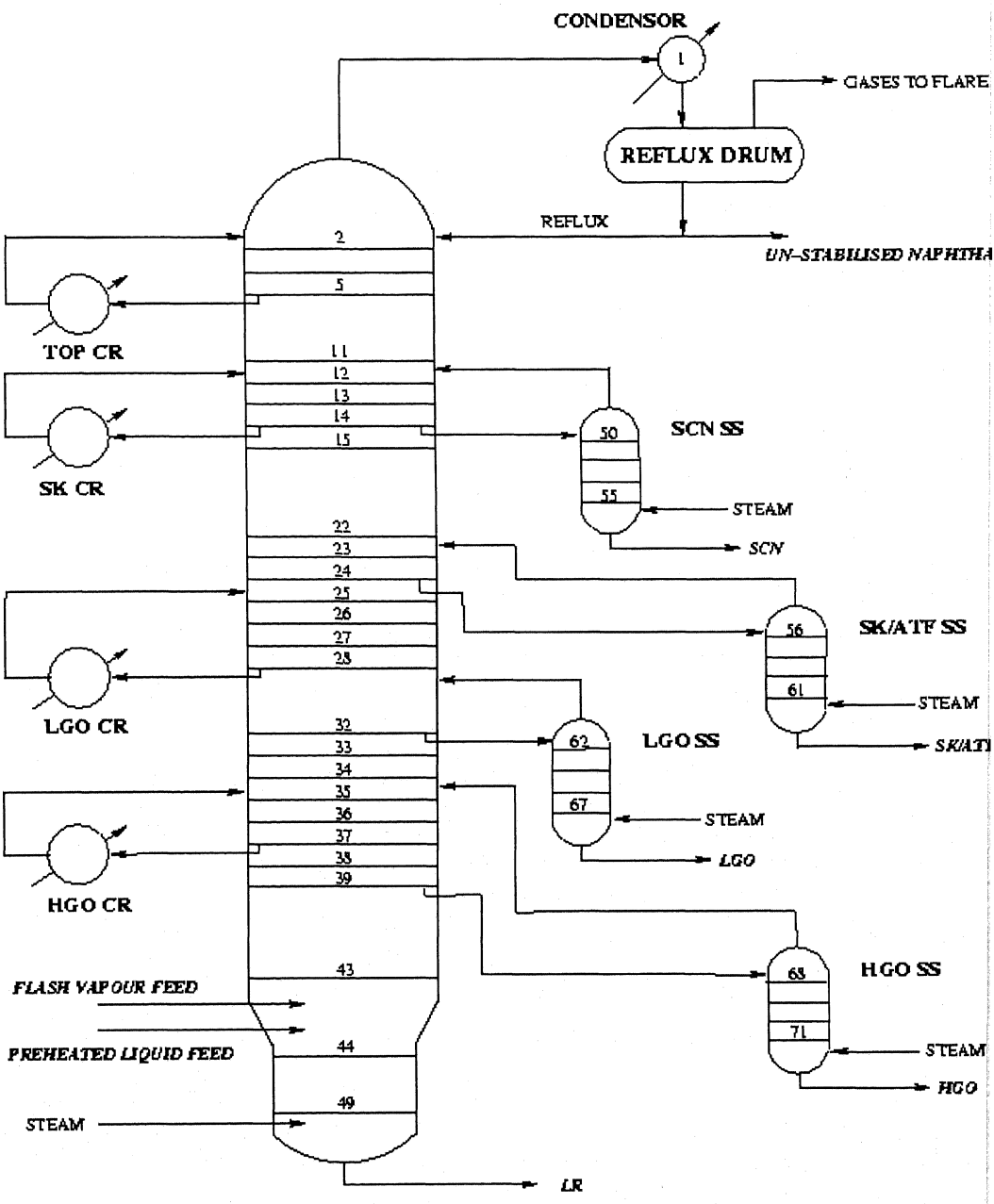


Fig. 1.2: SCHEMATIC DIAGRAM OF THE CRUDE DISTILLATION COLUMN

being distillate product, is fed to a stabilizer unit. Several operating data, temperature, pressure and flow rates are measured online and recorded as a function of time. These data are used for process monitoring and control. However, because of presence of both random and gross error in these measured raw data, it is necessary that these data be corrected for both types of errors before further use.

Present Work:

An attempt is made to detect the gross error in measured data coming from the crude distillation unit. Generalized Likelihood Ratio (GLR) method (Narsimhan and Mah, 1987) is used for detection of the gross errors. There are two types of measurements in the CDU, namely, flow rates and temperatures. Though much has been said about the gross error detection in flow rates, nobody has reported the gross error detection in temperatures. In the present work, we have tried to detect the gross errors in both, flow rates and temperatures. Chapter 2 includes a literature review. The GLR method used for the gross error detection is discussed in Chapter 3. Estimation of covariance matrix from the process data required for gross error detection scheme is also discussed in Chapter 3. Chapter 4 presents formulation of gross error detection problem for steady state CDU data. An attempt is made to solve the problem using the algorithm based on GLR technique. Chapter 5 presents results and discussion. The last chapter includes conclusions drawn from the present work and some recommendations made for future study.

Chapter 2

LITERATURE REVIEW

In practice the raw process data may contain gross errors and random errors. While the former are caused by non-random events such as leaks, deposition and inadequate accounting of departures from steady state operation as well as measurement biases and malfunctioning instruments, the latter may arise because of process noise, turbulence around the sensor etc. A lot of research has been done in the field of gross error detection and several statistics tests have been proposed for gross error detection, based on the assumption that the measurements are a random sample of the true values at the steady state. A brief summary of work done is presented below in the chronological order. Since data reconciliation and gross error detection is a twin problem and needs to be tackled together, most of the researchers have worked on both data reconciliation and gross error detection together and that is why we review the literature simultaneously.

- Mah et al. (1976) showed how information inherent in the process constraints and measurement statistics can be used to reconcile the data. Two graph-theoretic results were derived and used to simplify the reconciliation and the estimation of unmeasured process streams. The scheme was implemented and evaluated for a 32 node, 61 stream problem, and the results indicated 60% reduction in total absolute errors.
- Mah and Tamhane (1982) discussed Measurement Test to cope up with the problem of detecting and identifying the presence of one or more gross errors in the process data.
- Almasy and Mah (1984) outlined a method for estimating the error variances i.e. covariance matrix from the available process data. They proposed both direct and indirect methods for calculating the covariance matrix. The fact that the covariance matrix can be estimated and updated as time proceeds leads to on-line gross error detection.
- Mah and Tamhane (1985) reviewed, in detail, most of the methodologies available at that time for data reconciliation and gross error detection. Several example problems were discussed along with their practical and theoretical limitations. The gross error detection methods discussed by these authors included Global Test,

Nodal Test and Measurement Test. They recommended Measurement Test because of its superiority over others in terms of identification of the gross error.

- Serth and Heenan (1986) presented several algorithms for the detection of gross errors in the process data on a steam metering system. They compared the performance of these algorithms, which included Iterative Measurement Test (IMT), Modified Method of Pseudonodes (MMP), Modified Iterative Measurement Test (MIMT). They also compared their algorithms with those of Measurement Test (MT) and Method of Pseudonodes and found that MIMT algorithm performed better than other methods for the system under consideration.
- Mah and Rosenberg (1987) proposed two new composite tests, Dynamic Measurement Test (DMT) and Extended Measurement Test (EMT) for gross error detection. These methods can be used to reduce mispredictions of identifying gross errors and they also account for bound violation of measurements. These tests can be carried out when more than one gross error is present.
- Narsimhan and Mah (1987) described the application of a new statistical approach based on Generalized Likelihood Ratio (GLR) for the gross error detection. Willsky and Jones (1974) originally adopted this approach to identify abrupt failures in dynamic systems. In addition to the identification of gross errors, this method estimates the magnitude of gross error, which is useful for judging the impact on the process. It was also shown that this approach is better than all other statistical tests as it is based on serial compensation while all other approaches are based upon serial elimination. Serial compensation strategy is more useful when one has to detect multiple gross errors. In this strategy one gross error at a time is located and compensated for it using its estimated magnitude before identifying more gross errors.
- Biegler and Tjoa (1991) presented a simultaneous strategy for data reconciliation and gross error detection. In this study instead of minimizing the error function in the least square sense, they formed a function which takes into account both contributions from random and gross errors. The effectiveness of this strategy is demonstrated on non-linear example problems.
- Crowe et al. (1992) presented the Maximum Power (MP) test which allows direct detection of gross errors in species balances around individual process units. It is

shown that the square of MP test statistic is precisely equal to the GLR test. The tests are illustrated with examples.

- Crowe and Tong (1995) presented a new approach based on Principal Component Analysis (PCA) for gross error detection and compared it with other techniques. It is shown that the new test is capable of detecting gross errors of small magnitude where the other tests fail and has substantial power to correctly identify the variables in error.
- Yang et al. (1995) reviewed all the existing methods of gross error detection at that time. They concluded that no method has a guarantee of consistently finding all the gross errors in the process data. A combinatory method, based upon measurement test and nodal test, has been proposed by these workers for practical implementation.
- Ramagnoli et al. (1996) described a strategy that allowed the identification of gross error for pyrolysis reactor measurements. A reactor model was formulated in terms of heat and mass balance equations and a GLR technique was used for gross error detection. A neural network trained with a robust back propagation algorithm, relating variables in the convective zone, was also used for gross error detection and results compared with those from GLR method.
- Kim et al. (1997) presented the Modified Iterative Measurement Test (MIMT) for gross error detection using non-linear programming techniques to improve its robustness and performance. The algorithm when tested on a CSTR example showed improved robustness compared to other gross error detection algorithms. This approach is useful for highly non-linear chemical processes.
- Jiang et al. (1998) presented a method to identify and estimate gross errors in linear dynamic data of a plant. It allowed multiple gross error detection. A serial compensation strategy was found applicable for identification and estimation of multiple gross errors.

Observations:

- Several methods discussed by the workers listed above require data reconciliation to be carried out before undertaking gross error detection. These methods include Measurement Test, Global Test and Nodal Test.

- Most of the references cited above presented examples from laboratory measurements, where reconciliation and/or gross error detection was carried out but only in mass or volume flow (using material balance constraints). No study has, however, been reported in published literature where gross error detection was attempted in temperature measurements.

Chapter 3

GENERALIZED LIKELIHOOD RATIO METHOD FOR GROSS ERROR DETECTION

Gross error detection in steady state chemical processes has received considerable attention in the last two decades and several statistical tests have been developed (Tamhane and Mah 1985) for this purpose. All methods for gross error identification make use of one or more statistical tests in combination with an identification strategy. These methods can be categorized as follows (Yang et al., 1995)

a) Based on the assumption that the data follow a normal distribution:

i) Global Test
ii) Maximum Power Test
iii) Constraint or Nodal Test
iv) Measurement Test
v) Generalized Likelihood Ratios
vi) Principal Component Analysis

b) Based on data abnormal distribution:

i) Bivariate Likelihood Distribution
ii) Non-central Probability Density Function

c) Neural Network Methods

None of the methods of gross error detection developed so far can guarantee of consistently finding all of the errors in any measured data set. Generalized Likelihood Ratio (GLR) method was chosen for the present study in preference to others because of its supposed superiority in gross error detection. This method was used for identifying gross errors in flow and temperature measurements in a crude distillation unit when in steady state operation. GLR method was found better than other available methods because in addition to its generality, we also obtain an estimate of the magnitude of the gross error, which is useful in judging the impact of the gross error on the process. GLR was found more useful particularly in the case of multiple gross errors. In order to detect and identify multiple gross errors, a serial elimination strategy is often used (Serth and Heenan, 1986). Typically, in the serial elimination strategy

measurements suspected of containing gross error are eliminated, the test statistic is recomputed, and if its value is now below the critical value, then the suspected measurements are declared in gross error. GLR uses serial compensation strategy for detection of multiple gross errors. In this strategy one gross error is identified at a time and compensated for it using its estimated magnitude before detecting more gross errors.

Process Model

The steady state model of a chemical process in the absence of gross error can be described by

$$z = x + v \tag{3.1}$$

$$Ax = 0 \tag{3.2}$$

Equation 3.1 is the measurement model where z: n x 1 is a measurement vector, x: n x 1 is a vector of true values of state variables and v: n x 1 is a vector of measurement errors. The measured errors are assumed to be normally distributed with zero mean and known covariance matrix Q. Equation 3.2 describes the linear or linearized mass and energy constraint matrix.

Measurement bias model

The model for a bias of unknown magnitude b in measuring instrument i is given by

$$z = x + v + be_i \tag{3.3}$$

where e_i is a vector with unity in position i and zero elsewhere.

The GLR Method:

Procedure for single gross error

Given measurements, z, we would like to determine the presence of each gross error, if any is present, and estimate its magnitude. We shall first consider the case when at

most one gross error is present. If the measurement has the error, the linear constraint $\boldsymbol{Az}$ would not be equal to zero.

Let the balance residual be given by

$$\boldsymbol{r} = \boldsymbol{Az} \tag{3.4}$$

$\boldsymbol{r}$: n x 1 is the residual vector.

If no gross errors are present, we can show that

$$E[\boldsymbol{r}] = 0 \tag{3.5}$$

$$Cov[\boldsymbol{r}] = \boldsymbol{V} = \boldsymbol{AQA}^T \tag{3.6}$$

where $E[\boldsymbol{r}]$ is expected value of $\boldsymbol{r}$ and $Cov[\boldsymbol{r}]$ is covariance matrix of $\boldsymbol{r}$. Estimation of $\boldsymbol{Q}$ matrix will be discussed in the next section.

If a gross error due to bias of magnitude b is present in measurement i, then by using Eqs. 3.2, 3.3, and 3.4 we can show that

$$E[\boldsymbol{r}] = b\boldsymbol{Ae}_i \tag{3.7}$$

Therefore, if we define μ as the unknown expected value of $\boldsymbol{r}$, we can formulate the hypothesis for gross error detection as (Narsimhan and Mah, 1987)

$$H_0 : \mu = 0 \tag{3.8}$$

$$H_1 : \mu = b\boldsymbol{Ae}_i \tag{3.9}$$

where H_0 is the null hypothesis that no gross errors are present and H_1 is the alternative hypothesis that a gross error is present. The alternative hypothesis has two unknown parameters, b and $\boldsymbol{Ae}_i$. The parameter b can be any real number and $\boldsymbol{Ae}_i$ can be any vector from the set $\boldsymbol{F}$, which is given by

$$\boldsymbol{F} = \{\boldsymbol{Ae}_i, : i = 1.....n\} \tag{3.10}$$

In order to test the hypothesis given by equation 3.8 and 3.9 and estimate of unknown parameters, if H_1 is accepted, we make use of the likelihood ratio test. The likelihood ratio test statistic in our case is given by

$$\lambda = \sup \frac{\Pr\{\boldsymbol{r} \mid H_1\}}{\Pr\{\boldsymbol{r} \mid H_0\}} \tag{3.11}$$

where the supremum in equation 3.11 is computed over all possible values of the parameters present in the hypothesis. Using the normal probability density function for $\boldsymbol{r}$, we can write equation 3.11 as

$$\lambda = \sup_{b, \boldsymbol{Ae}_i} \frac{\exp[-0.5\{\boldsymbol{r} - b(\boldsymbol{Ae}_i)\}^T \boldsymbol{V}^{-1} \{\boldsymbol{r} - b(\boldsymbol{Ae}_i)\}]}{\exp[-0.5 \boldsymbol{r}^T \boldsymbol{V}^{-1} \boldsymbol{r}]} \tag{3.12}$$

Since the expression on the right hand side of equation 3.12 is always positive, we can simplify the calculation by choosing as the test statistic

$$T = 2 \ln \lambda = \sup_{b, \boldsymbol{Ae}_i} [\boldsymbol{r}^T \boldsymbol{V}^{-1} \boldsymbol{r} - \{\boldsymbol{r} - b(\boldsymbol{Ae}_i)\}^T \boldsymbol{V}^{-1} \{\boldsymbol{r} - b(\boldsymbol{Ae}_i)\}] \tag{3.13}$$

The computation of T proceeds as follows. For any vector $\boldsymbol{Ae}$, we compute the estimated magnitude of b i.e. $\hat{b}$ which gives the supremum in equation 3.13. Thus we obtain the maximum likelihood estimate

$$\hat{b} = \{(\boldsymbol{Ae}_i)^T \boldsymbol{V}^{-1} (\boldsymbol{Ae}_i)\}^{-1} \{(\boldsymbol{Ae}_i)^T \boldsymbol{V}^{-1} \boldsymbol{r}\} \tag{3.14}$$

Substituting this value of $\hat{b}$ for b in equation 3.13 and denoting the corresponding value of T by T_i we get

$$T_i = d_i^2 / C_i \tag{3.15}$$

where

$$d_i = (\boldsymbol{A}\boldsymbol{e}_i)^T \boldsymbol{V}^{-1} \boldsymbol{r} \tag{3.16}$$

$$C_i = (\boldsymbol{A}\boldsymbol{e}_i)^T \boldsymbol{V}^{-1} (\boldsymbol{A}\boldsymbol{e}_i) \tag{3.17}$$

This calculation is performed for every vector $\boldsymbol{A}\boldsymbol{e}_i$ in set $\boldsymbol{F}$ and the test statistic T is therefore obtained as

$$T = \sup_i T_i \tag{3.18}$$

Let $\boldsymbol{f}^*$ be the vector that leads to the supremum in equation 3.18. The test statistic T is compared with a prespecified threshold ε, and a gross error is detected if T exceeds ε. The value of ε is got from the normal probability density table for 5% level of significance and 1 degree of freedom. The gross error that corresponds to the vector $\boldsymbol{f}^*$ is identified as the gross error and its magnitude is estimated from equation 3.14 using $\boldsymbol{f}^*$ for $\boldsymbol{f}_i$.

Strategy for multiple gross errors

The above treatment pertains to the detection and identification of a single gross error. For multiple gross errors a strategy based on serial compensation is used. In this strategy, we identify one gross error at a time by applying the GLR test. If a gross error is identified, we use its estimated magnitude to compensate for the gross error. This process is repeated until no further gross errors are detected. The algorithm for implementing the serial compensation strategy with the GLR test is outlined in Fig. 3.1. Strictly speaking, the serial procedure described above works better if the detection of a gross error is not affected by the presence of any other gross error.

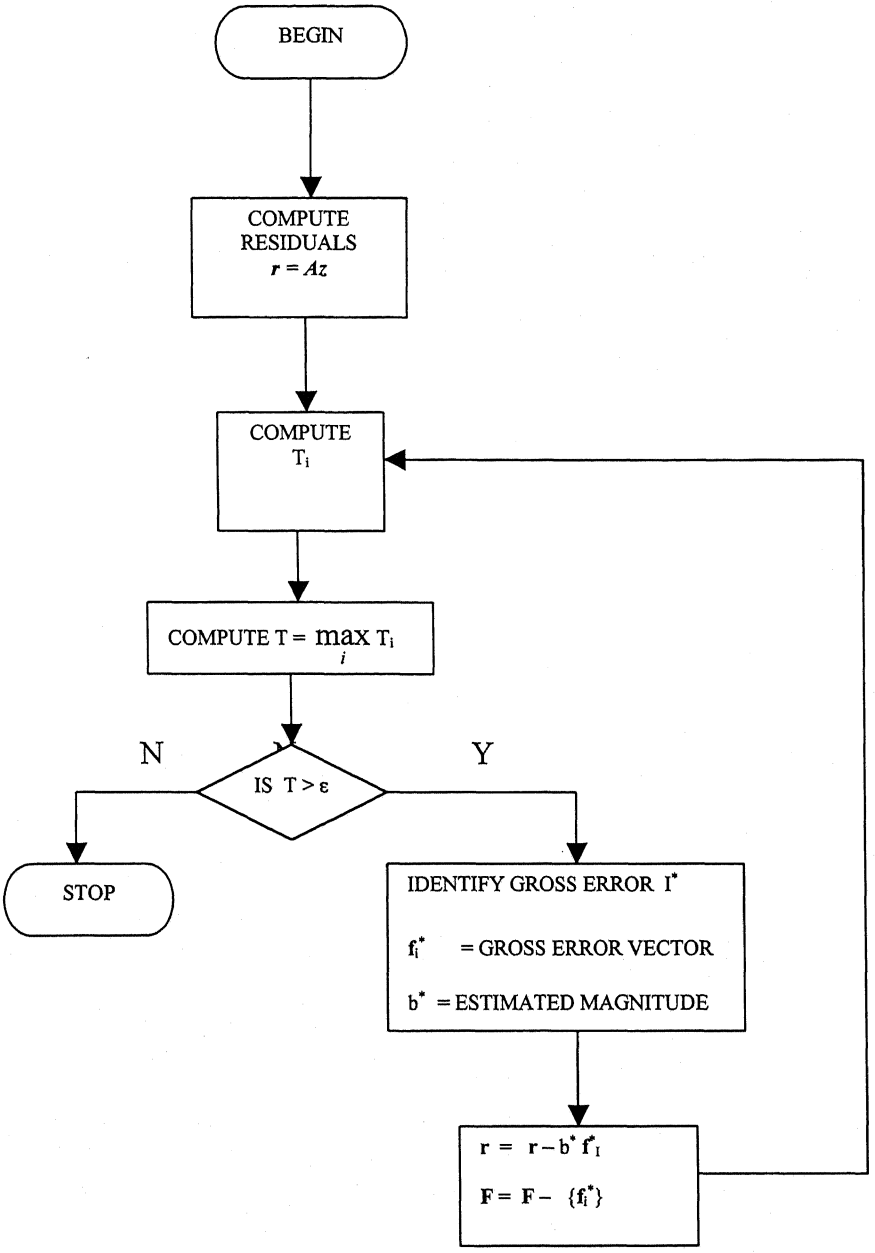


Figure 3.1: GLR test with serial compensation strategy

Covariance Matrix Estimation

Any gross error detection method requires a good estimation of covariance matrix, $\boldsymbol{Q}$. Almost all systematic techniques start with the assumption of a known covariance matrix of measurement errors. However, little has been said on how such knowledge might have been obtained in the first place. (Almasy and Mah, 1984) described methods of estimating $\boldsymbol{Q}$ for real processes where the variations in the measured variables in different sets of the measurements are comparable in magnitude. Some details of the method for calculation of the $\boldsymbol{Q}$ matrix are given below.

A. Direct Method

Let z_i be the measurement of i^{th} variable in the measured variable set. Then sample variance of m repeated measurements of z_i is given by

$$\text{var}(z_i) = \frac{1}{m-1}\sum_{k=1}^{m}(z_{ik} - \overline{z}_i)^2 \tag{3.19}$$

And the covariance of z_i and z_j is given by

$$\text{var}(z_i) = \frac{1}{m-1}\sum_{k=1}^{m}(z_{ik} - \overline{z}_i)(z_{jk} - \overline{z}_j) \tag{3.20}$$

where $\overline{z}_i$ is the average of i^{th} variable over m data sets

In the $\boldsymbol{Q}$ matrix, diagonal terms are the variance terms and off-diagonal terms are the covariance terms.

B. Indirect Method

Rewriting conservation constraints i.e. equation (3.2) typically take the form of

$$\boldsymbol{Ax} = 0 \tag{3.2}$$

In general equation (3.2) will not be satisfied by the measurement values of variables z_i. Let the residuals of the constraints be denoted by $\boldsymbol{r}$.

$$\boldsymbol{r} = \boldsymbol{Az} \tag{3.4}$$

Under H_0 hypothesis the expected value of $\boldsymbol{r}$,

$$E(\boldsymbol{r}) = E(\boldsymbol{A}\boldsymbol{z}) = 0 \tag{3.5}$$

$$\begin{aligned} \text{cov}(\boldsymbol{r}) &= E([\boldsymbol{r} - E(\boldsymbol{r})][\boldsymbol{r} - E(\boldsymbol{r})]^T) \\ &= \boldsymbol{A}\boldsymbol{Q}\boldsymbol{A}^T \\ &= \boldsymbol{H}(say) \end{aligned} \tag{3.21}$$

Residual variance covariance matrix is directly calculated from the measured data without requiring the true values of the variables. The calculation of $\boldsymbol{H}$ from the residuals allows us to back out the estimate of $\boldsymbol{Q}$. Now for given $\boldsymbol{Q}$ and $\boldsymbol{A}$, $\boldsymbol{H}$ is uniquely determined, but the converse is not true. It is observed that since the measurements of different variables are normally made independently, the measurement errors should be uncorrelated. To allow for the possibilities that they may share some common elements such as power supplies, we shall assume that they are weakly correlated. The implication is that $\boldsymbol{Q}$ will be diagonal or diagonally dominant. One basis to estimate $\boldsymbol{Q}$ is to minimize the sum of squares of the off-diagonal elements of $\boldsymbol{Q}$ subject to the constraint

$$\boldsymbol{H} = \boldsymbol{A}\boldsymbol{Q}\boldsymbol{A}^T \tag{3.22}$$

The program code given by Almasy and Mah (1984) for the calculation of $\boldsymbol{Q}$ although worked well for the trial case, reported by them, but in absence of any details, it was not possible to use their procedure in the present study. The problem was, therefore, formulated as follows and the optimization was carried out using Sequential Quadratic Programming (Dutta, 1999).

Let the off-diagonal elements of $\boldsymbol{Q}$ be q_{ij} and diagonal elements are q_{ii}

$$\text{Objective Function} = \sum_{i \neq j} q_{ij}^2 \tag{3.23}$$

The above objective function is minimized subject to

$$\boldsymbol{A}\boldsymbol{Q}\boldsymbol{A}^T = \boldsymbol{H} \tag{3.24}$$

An Example Problem

Process network shown in Figure 3.2 is taken as example problem to test the GLR method. This example is taken from Mah, (1990). All five streams are measured. This system has a high degree of redundancy. A feed stream containing nitrogen (N_2) and hydrogen (H_2) is mixed with the recycle stream 5 at node 'a'. The reaction according to equation (3.25) takes place in the reactor (node 'b'). The product, ammonia (NH_3), is recovered in the separator (node 'c'), and the unreacted gases are recycled to the node 'a'.

$$N_2 + 3H_2 \longrightarrow 2NH_3 \tag{3.25}$$

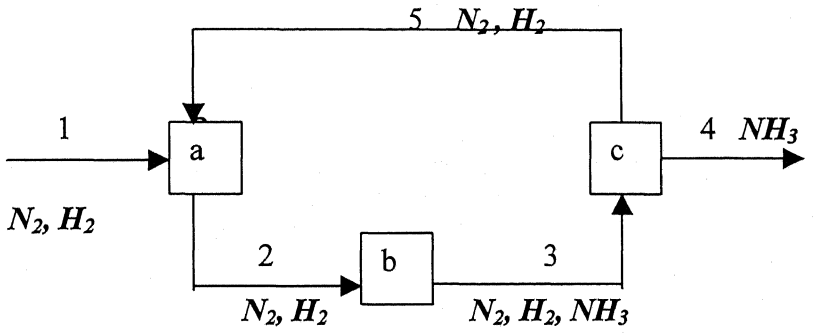


Figure 3.2 Process Network for ammonia synthesis

The incidence matrix of the above network is given below

$$A = \begin{bmatrix} 1 & -1 & 0 & 0 & 1 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & -1 \end{bmatrix} \tag{3.26}$$

True value of the measurements are given as

$$z = [225 \quad 325 \quad 325 \quad 225 \quad 100]^T \tag{3.27}$$

Fifty different measurement sets were generated based on this true value set. These

fifty sets were generated using a subroutine (RNNOA: IMSL) for random number generation. This subroutine generates the random number with mean equal to zero and variance equal to one. Then with the help of these fifty data sets (see Appendix A) we generated the $\boldsymbol{Q}$ matrix (using direct method). The $\boldsymbol{Q}$ matrix obtained is shown below.

$$\boldsymbol{Q} = \begin{bmatrix} 1.061 & 0.054 & 0.087 & -0.020 & 0.014 \\ 0.054 & 1.142 & -0.092 & 0.501 & 0.000 \\ 0.087 & -0.094 & 1.115 & 0.078 & -0.057 \\ -0.020 & 0.501 & 0.078 & 1.015 & 0.057 \\ 0.014 & 0.000 & -0.057 & 0.057 & 1.125 \end{bmatrix} \tag{3.28}$$

We perturbed the measurements with some known gross errors (one measurement at a time) and tried to find that value by running the code written on the basis of GLR method.

The program was able to detect the error with its correct magnitude. Some of the results obtained are shown below.

Table 3.1 Results of GLR method applied to an example problem

Sr. no.	Error in stream no.	Magnitude of error introduced	Magnitude of error detected
1.	1	6.5	6.05
2.	2	8.4	8.1
3.	4	5.7	5.3

Chapter 4

MATHEMATICAL FORMULATION

Before proceeding with the detection of the gross errors in the measurements on the crude distillation unit it is necessary to represent the unit in terms of a network called diagraph (see Apendix B for a brief review of graph theory). The CDU under consideration has 49 stages in the main column and, therefore, the diagraph for this unit will have 49 nodes. The total number of liquid and vapor streams involved is 99, which connects these nodes . However, in the absence of sufficient measurements being made, it is not possible to proceed with this network. Actually there are only 11 stages in the column where the flows and/or temperatures are measured, which necessitates the network to contain only as many nodes. The reduced network is shown in Figure 4.1. In the figure, the number inside the circle represents the number of the actual plate inside the column. For example the number 14 inside the fourth node represents the 14^{th} plate of the actual column. The downward arrows in the figure show the flow of liquid streams while the upward arrows show the flow of vapor streams. The streams in the right side of the figure show the products withdrawn from the column through the side-strippers. The streams in the left of the figure show the pump-around flows.

In this reduced network, twenty-three variables are measured. These 23 measurements can be divided into two categories, namely, flow rates and temperatures. These measured variables are as follows:

Flow rates of

1. un-stabilized naphtha (UN)
2. reflux
3. top circulation reflux (CR)
4. superior kerosene (SK) CR
5. straight cut naphtha (SCN) draw
6. superior kerosene (SK) draw
7. light gas oil (LGO) CR
8. LGO draw

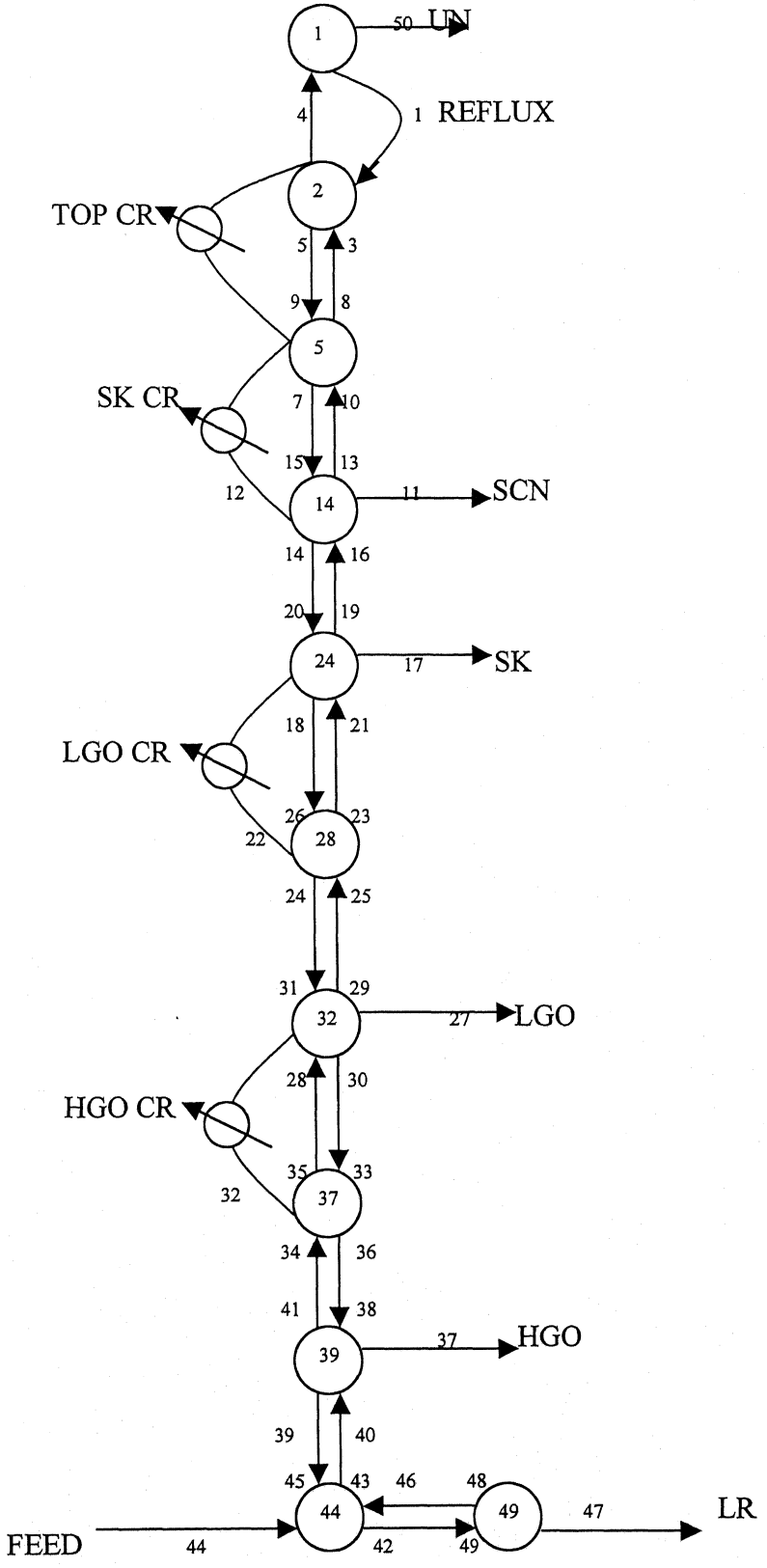


Fig. 4.1 Reduced Network for the Crude Distillation Unit

9. heavy gas oil (HGO) CR
10. HGO draw
11. liquid feed
12. long residue (LR)

Temperatures of

1. condenser
2. column top
3. top CR draw
4. SK CR draw*
5. SCN draw*
6. SK draw
7. LGO CR draw
8. LGO draw
9. HGO CR draw
10. HGO draw
11. feed
12. column bottom

**SCN draw and SK CR draw are from the same plate so SCN draw temperature will be the same as SCN draw temperature and thus we know only eleven temperatures in place of twelve*

As we can see in Figure 4.1, there are 31 streams in the process diagraph. But in actual case there will be more than 31 streams. For example, the liquid stream leaving stage 2 will not be the same as the liquid stream entering stage 5 since in actual case stages 2 and 5 are not be adjacent to each other, but there are 2 more plates between these. Likewise the vapor stream leaving stage 5 will not be the same as the vapor stream entering stage 2. Because of this reason we take four streams to be present between the nodes 2 and 5 namely liquid leaving stage 2, vapor leaving stage 5, liquid leaving stage 4 (which is same as to liquid entering stage 5) and vapor leaving stage 3 (which is same as the vapor entering stage 2). When we separate the streams like this between each two nodes, there will be a total of 50 streams in this process diagraph. These fifty streams are listed below:

1. reflux
2. top CR pump around return
3. vapor leaving plate 3
4. vapor leaving plate 2
5. liquid leaving plate 2
6. top CR pump around draw from plate 5
7. liquid leaving plate 5
8. vapor leaving plate 5
9. liquid leaving plate 4
10. vapor leaving plate 6
11. SCN draw from plate 14
12. SK CR pump around draw from plate 14
13. vapor leaving plate 14
14. liquid leaving plate 14
15. liquid leaving plate 13
16. vapor leaving plate 15
17. SK draw from plate 24
18. liquid leaving plate 24
19. vapor leaving plate 24
20. liquid leaving plate 23
21. vapor leaving plate 25
22. LGO CR pump around draw from plate 28
23. vapor leaving plate 28
24. liquid leaving plate 28
25. vapor leaving plate 29
26. liquid leaving plate 27
27. LGO draw from plate 32
28. liquid leaving plate 32
29. vapor leaving plate 32
30. vapor leaving plate 33
31. liquid leaving plate 31
32. HGO CR pump around draw from plate 37
33. vapor leaving plate 37
34. liquid leaving plate 37

35. liquid leaving plate 36
36. vapor leaving plate 38
37. HGO draw from plate 39
38. vapor leaving plate 39
39. liquid leaving plate 39
40. vapor leaving plate 40
41. liquid leaving plate 38
42. liquid leaving plate 44
43. vapor leaving plate 44
44. liquid feed
45. liquid leaving plate 43
46. vapor leaving plate 45
47. liquid leaving plate 49
48. vapor leaving plate 49
49. liquid leaving plate 48
50. un-stabilized naphtha

Gross Error Detection in Flow Rates

For gross error detection in flow rates we have to perform the material balance at each node of the process diagraph with the help of given flow rate measurements. But as discussed above, fifty different streams are involved in the process diagraph and out of these fifty streams only twelve are measured. With the help of these twelve measured-streams, it is not possible to perform material balance at each node. For a good estimation of material balance, we have to have an estimate of the flow rates of all the fifty streams. Unfortunately, it is not possible to get these data from an operating CDU, as no refinery measures the liquid and vapor flow rates inside the column. So, to have an estimate of these unmeasured streams, a simulation model of the CDU was used which was developed in-house (Prasanth, 2000). This simulator needs a few operating conditions to run which are available. This simulator can give us the liquid and vapor flow rates throughout the column.

With the available twelve flow rates that are measured and the estimated flow rates of the rest of the streams, material balance at each node can be performed. If we assume

that one or more flow rate measurements have gross errors, the process constraint equation $Az = 0$ will not hold good and the residual will be

$$r = Az \tag{4.1}$$

Now this residual vector **'r'** is passed through the statistics based on the GLR method (as described in Chapter 3) in order to detect the error.

Q matrix estimation

As described in Chapter 3, a good estimation of covariance matrix of measurement errors, Q, is a prerequisite for any gross error detection method to work. For Q matrix estimation, we should have a large number of data sets, say 100. These data sets can be assumed to be randomly distributed about the true values with mean zero and variance one. If we know one data set, which we can call as true set of values at a particular instant of time, we can generate any number of data sets with mean zero and covariance one, using the programs for random number generation. But this procedure requires the availability of a true data set. However, data collected online on the operating CDU at any particular time can not be taken as the true value. To overcome this difficulty, several (6 to 7) measured consecutive data sets were collected and the liquid and vapor flow rates were calculated using the simulator. Average values were then found for the measured and calculated flow rates over all the data sets. These average values were then taken as the true values of the measurements. Using a random number generator (RNNOA : IMSL), 100 data set were generated which were, then, used to calculate the Q matrix. The size of Q matrix is (50 × 50) as there are fifty measurements in the process diagraph.

Strategy for On-line Gross Error Detection in Flow Rates

The development of this package of gross error detection is aimed for on-line use. First of all a true data set is generated with the help of the simulator as discussed above. The estimated values of unmeasured flow rates are assigned to the corresponding streams of the process diagraph. The measured flow rates will come on-line. With the help of the estimated (unmeasured) and on-line (measured) flow rates, the GLR test is performed and the outlier and its magnitude detected as described in Chapter 3. However, the flow

However, the flow rates of some liquid and vapor streams which have been estimated through the simulator, need to be changed (updated) after a certain period of time or if and when there is any major change in the operating conditions or in crude type. However, it is safe practice to update these flows say, every hour or so even under normal operation. If there are more than one measured flow rates in error, we use the serial compensation strategy (Figure 3.2). In this case the measurement with the largest amount of error will be detected first and then it will be compensated by the estimated amount. Then the subsequent errors will be detected.

Gross Error Detection in Temperatures

Gross error detection in temperatures is very much different from that in flow rates. One major difference is that in this case we have to perform the enthalpy balance at each node but enthalpy is not a measured quantity. Enthalpy must be calculated using measured temperatures and some empirical correlations which are at best approximate and introduce errors of varying magnitudes for different components in different temperature ranges. The composition of petroleum fractions is truly an undefined parameter and must be expressed in terms of arbitrarily chosen pseudocomponents. All these difficulties coupled with scanty measured data make gross error detection in temperature, a truly difficult job. Several studies have been reported in published literature on gross error detection in flow rates, but nobody has reported the gross error detection in temperatures. We have made an attempt to detect the gross errors in the temperatures with the help of the limited available data and the CDU simulator. As there are fifty streams in the process diagraph, the enthalpies of all these fifty streams must be calculated or estimated. This required some assumptions to be made which are given below.

Assumptions

Kesler and Lee (1976) modification of Johnson-Grayson charts is applicable for estimation of liquid and vapor enthalpies of pseudo-components.

$$\text{Liquid enthalpy } h_j = \int_{200(^\circ F)}^{T(^\circ F)} (C_{PL})_j \, dT \tag{4.2}$$

$$\text{Vapour enthalpy } H_j = \int_{200(^\circ F)}^{T(^\circ F)} (C_{PV})_j \, dT + (\Delta H_V)_j \tag{4.3}$$

where,

$(C_{PL})_j$ = Heat capacity of the j^{th} component in liquid phase (BTU/lb-°F).

$(C_{PV})_j$ = Heat capacity of the j^{th} component in vapor phase (BTU/lb-°F).

$(\Delta H_V)_j$ = Heat of vaporisation of j^{th} component (BTU/lb).

The expression for *liquid heat capacity* is:

$$(C_{PL})_j = \left(0.6811 - 0.308(Sg)_j + \left[0.000815 + 0.000306(Sg)_j\right]T\right) \times \left(0.35 + 0.055K_w\right) \quad (4.4)$$

The molar enthalpy of the required liquid stream is calculated by adding the pseudo-component enthalpies in proportion to their mole-fractions in the mixture. The vapor molar enthalpy is also calculated similarly using the following expression for vapor heat capacity.

$$(C_{PV})_j = A + BT + CT^2 - C_F(A' + B'T + C'T^2) \quad (4.5)$$

where,

$$A = -0.32646 + 0.02678K_w$$

$$B = -(1.3892 - 1.2122K_w + 0.03803K_w^2)10^4$$

$$C = -1.5393X10^{-7}$$

$$A' = 0.084773 - 0.080809(Sg)_j$$

$$B' = -[2.1773 - 2.0826(Sg)_j]10^{-4}$$

$$C' = [0.78649 - 0.70423(Sg)_j]10^{-7}$$

$$C_F = \left[\left(\frac{12.8}{K_w} - 1\right)X\left(\frac{10}{K_w} - 1\right)X100\right]^2$$

In the above expressions, temperature T is °F, K_w is the Watson Characterisation factor and $(Sg)_j$ is the specific gravity of j^{th} pseudo-component.

Heat of vaporisation

Lee and Kesler (1976) calculated heat of vaporisation analytically based on Pitzer's three-parameter corresponding states principle. This method though accurate enough for compounds generally encountered in hydrocarbon processing, the calculation is rigorous and requires a significant amount of computation time. So, the correlation presented by Edmister and Lee (1984) for calculating heat of vaporisation has been used. However, a constant deviation was observed when compared with Lee-Kesler values. Hence, the equation was modified and a constant factor is added to the expression to get a satisfactory result. The expression for heat of vaporisation is given below.

$$(\Delta H_V)_j = (4.00495 + 5.456344\omega_j)R(T_c)_j\left(\frac{1.8}{M_j}\right) + C_1 \qquad (4.6)$$

where,

$(\Delta H_V)_j$	=	heat of vaporisation of the j^{th} component (BTU/lb).
R	=	universal gas constant (= 1.986 BTU/lbmole-°F)
$(T_c)_j$	=	critical temperature of the j^{th} component in Kelvin.
ω_j	=	acentric factor of the j^{th} component.
M_j	=	molecular weight of the j^{th} component.
C_1	=	a constant added for matching the result with that of Lee-Kesler Equations and is dependent on crude type.

To use these correlations we should know the composition of a hydrocarbon stream in terms of its pseudo-components. The CDU simulator (Prasanth, 2000) breaks the feed (crude oil) into 25 pseudo-components, and also calculates the compositions of all the intermediate streams in terms of these pseudo-components. The other properties of pseudo-components like specific gravity, molecular weight, critical temperature, acentric factor are calculated using empirical correlations. In other words, if we know the temperature of a stream (measured or estimated) and its composition in terms of pseudo components (from the simulator), we can get the specific enthalpy of that stream using the above correlations. This specific enthalpy when multiplied by the flow rate of the corresponding stream gives the enthalpy.

Now our task is reduced to get the composition of all the streams in terms of pseudo components and temperatures or specific enthalpies of some of the streams (in particular the streams that are coming to the 11 nodes). This problem was solved by using the CDU simulator. Several sets of data were collected on-line and various stream compositions calculated using the simulator. It was found that the composition of the streams did not differ much from one set to another as long as operating conditions did not change significantly. With this conclusion, the compositions of the streams in the process diagraph were temporarily fixed. Now the only thing required for specific enthalpy calculation was the temperature of the streams. It has been pointed out earlier that the temperatures are measured only at the eleven locations (eleven nodes) inside the column which correspond to temperatures of the streams leaving these eleven nodes e.g. streams 4, 5, 11, 13, 14, 32, 33, 34 etc. But the temperatures of the streams which are coming to these nodes are not available e.g. streams 3, 9, 10, 25, 26, 35, 36 etc. To overcome this problem, the values of specific enthalpies of these streams for a number of data sets were examined. It was found that if the operating conditions do not change much, the value of specific enthalpy also does not change significantly. This prompted us to temporarily fix the specific enthalpy of these streams. However, it should be pointed out that the specific enthalpy of these streams need to be changed (updated) if there is any major changes in the operating conditions or in the crude type or at regular intervals of time.

The above procedure allowed us to estimate the specific enthalpy of all the fifty streams. The total enthalpies of all the streams can be calculated by multiplying the specific enthalpies with the flow rates of respective streams. These flow rates are either measured or estimated with the help of CDU simulator.

Q matrix estimation

For Q matrix estimation we should have an estimate of the true values of enthalpies of the streams. For this purpose 6 to 7 consecutive on-line data sets were collected and the enthalpy of all the streams were calculated for all the sets as described above. The average values of these runs were taken as the true value of the enthalpies. Fifty different data sets were then generated using random number generator

(RNNOA:IMSL). These fifty data sets were used to get the $\boldsymbol{Q}$ matrix as described earlier.

Strategy for On-line Gross Error Detection in Temperatures

As discussed earlier, error detection in temperature measurements is dependent on enthalpy balance at each node but enthalpy is not a measured quantity. Moreover, since the enthalpy of a stream is proportional to its flow rate, it is important that gross error detection is first undertaken in flow rates and any error if present is compensated. To use the package for on-line purpose, a true set of temperature data is found as discussed earlier for flow rates. This set is used with the CDU simulator to calculate unmeasured flows, the composition and the specific enthalpies of the streams as per our requirements.

The calculated intermediate flow rates along with twelve measured flow rates and eleven measured temperatures were used to get the enthalpies of all the streams. It is now possible to perform the enthalpy balance at each node and the residuals sent through the GLR test in order to detect the presence of gross error in measured temperatures. The GLR test is expected to detect the node and the corresponding temperature that is in error. But since the method relies on enthalpy balance, instead of identifying the outlier temperature, it will estimate the magnitude of the discrepancy in the enthalpy due to the error in temperature. It is obvious that if the temperature of a particular node is in error, the enthalpies of all the streams leaving that node will be affected. And the discrepancy in enthalpy estimated by the GLR method at a particular node will be the sum of the discrepancies of all streams leaving that node. It was, therefore, necessary to distribute the enthalpy among all the streams leaving that node in proportion to their flow rates. Enthalpy of each stream leaving that node was corrected using its share of discrepancy. This will give us the estimate of the correct enthalpies of the streams leaving that node. Now we can back calculate the temperature that will give us the correct enthalpy for that stream. This is the corrected temperature which must be used for further computations.

If more than one temperature are in error, the temperature that will cause the maximum amount of discrepancy at the node, will be detected first. It must be compensated by its

estimated amount of discrepancy in enthalpy before proceeding to detect the second error in subsequent temperature measurements.

Strategy for on-line gross error detection in both, flow rates and temperatures

If the measured flow rates and temperatures both are in error, we will first detect the gross error in flow rates. After detecting the errors in all the flow rates and estimating their respective magnitude, these streams are compensated and then used in the simulator. The corrected measured flow rates and measured temperatures are then used to detect gross error in the temperatures.

Chapter 5
RESULTS AND DISCUSSION

The methodology for on-line gross error detection in the steady state data of CDU and the way it has been implemented have already been discussed in detail in the previous chapters. The results that have been obtained on a few typical cases are reported in this chapter.

Single error in flow rate measurements

In the first case the values of measured flow rates were perturbed by some known quantity one at a time and the algorithm checked for its ability to detect the gross error in the measurement and for its estimated compensation. The results obtained for a few typical cases are shown in Table 5.1. Only one flow rate was perturbed at a time either in +ve or in –ve direction. The third column in Table 5.1 gives the smallest magnitude of error that was detectable. The magnitude of most of these flow rates are around 1000 MT/day. So, we can say that in most of the cases, the algorithm is able to detect the errors to the tune of 1%. The measurement of flow rates is available in MT/day while the algorithm performs the material balance in lb-moles/hr. For this, MT/day flow rates are first converted into lb-moles/hr. Molecular weight of the streams increases as we move towards the bottom end of the distillation column. Streams like liquid feed, LR flow and HGO flow have a high molecular weight that reduces the amount of error in terms of lb-mole/hr. That is why the error of even 15-16 MT/day was not detectable in these streams, thus reducing the sensitivity of this algorithm.

Problem arises when two measured streams are connected to the same node. Consider the case of flow rates of SK CR and SCN draw streams. Both of these streams are connected to the same node (node 4, see Figure 4.1). So, if material balance does not match at node 4, it would be difficult to judge which stream is in error. In this case the detected error may be in SK CR draw or in SCN draw or in both. It would also be difficult in this case, to estimate the magnitude of error. We calculated the estimated amount of error assuming the error in both the streams one by one.

Table 5.1 Single error in flow rate measurements

Case no.	Flow rate perturbed	Magnitude of error introduced (MT/day)		Result obtained: Error detected in	Magnitude of estimated error	
		+ve	-ve		+ve	-ve
1	UN flow rate	10	11	UN flow	10.2	11.4
2	Reflux flow rate	9	9	Reflux flow	11.3	8.6
3	Top CR flow rate	10	10	Top CR flow	10.6	9.3
4	SK CR flow rate	11	10	SK CR or SCN flow	10.3 or 6.4	10.3 or 6.4
5	SCN flow rate	8	8	SK CR or SCN flow	14.1 or 8.3	13.6 or 7.8
6	SK flow rate	10	10	SK flow	9.7	10.4
7	LGO CR flow rate	10	10	LGO CR flow	9.2	10.7
8	LGO flow rate	10	10	LGO flow	11.2	11.2
9	HGO CR flow rate	10	10	HGO CR flow	11.2	11.2
10	HGO flow	14	14	HGO flow	13.2	14.7
11	Liquid feed flow rate	16	16	Liquid feed flow	18.2	18.5
12	LR flow rate	17	17	LR flow	18.2	18.5

If SCN draw was in error, the discrepancy in material balance was multiplied by the molecular weight of SCN and if SK CR pump around was in error, the discrepancy was multiplied by the molecular weight of SK CR, and both results are reported in the last two columns of the above table.

The other problem encountered was with the streams like reflux and Top CR each of which are connected with two nodes. If such a stream is in error, it will affect the material balance at two nodes. For example if reflux flow rate is in error it will affect the material balance at node 1 as well as at node 2. Then at node 1, the discrepancy may be due to reflux flow or due to UN flow and at node 2 it may be due to reflux flow or due to Top CR flow (see Figure 4.1). In this case first we temporarily compensated for UN flow and saw that the material balance at node 1 has matched but it hasn't matched at node 2. Then we compensated for reflux flow and found that the material balance at

both the nodes has matched. So the conclusion was drawn that actually reflux flow rate was in error. It is important here to undo the temporary compensation provided in UN flow once the source of error is identified. Similarly the conclusion for Top CR draw flow rate can also be drawn.

Two errors in flow rates measurements

Any two flow rate measurements were simultaneously perturbed with some known quantities and the algorithm was tested for its ability to detect multiple gross errors. The results obtained for a few typical cases are shown in Table 5.2.

Table 5.2 Two errors in flow rate measurements

Case No.	Flow rates perturbed simultaneously	Magnitude of error introduced (MT/day)	Result obtained: error (in order of detection) in	Magnitude of error estimated (MT/day)
1	i) SK flow ii) HGO CR flow	+ 13 + 17	i) SK flow ii) HGO CR flow	+ 11.2 + 19.4
2	i) SK flow ii)Liquid feed flow	+ 13 + 18	i) SK flow ii)Liquid feed flow	+ 15.1 + 20.1
3	i)Top CR flow ii) LR flow	+ 10 + 20	i) Top CR flow ii) LR flow	+ 12.2 + 17.7
4	i) Reflux flow ii) LGO CR flow	+12 + 13	i) Reflux flow ii) LGO CR flow	+ 13.6 + 11.2
5	i)SK flow ii) LGO CR flow	-14 -14	i) SK flow ii) LGO CR flow	-16.1 -12.2
6	i) SCN flow ii) Liquid feed flow	-11 -18	i) SK CR flow or SCN flow or both ii) Liquid feed	-18.0 or -12.8 -15.9

From Table 5.2 it is seen that, the detectable value of error has increased in presence of two errors. For example if SK draw flow rate alone was in error, the discrepancy of 10 MT/day was detectable whereas in this case it increased up to 14 MT/day. Moreover, the estimated amount of error also deviates from the actual amount of error more than reported in Table 5.1. This is a limitation of GLR method. As the number of error

increases, the detectability and the correctness of the estimated magnitude decreases. The other noticeable thing here is, though the magnitude of error introduced in the liquid feed flow rate is larger than that in the SK draw flow rate (Case 2), the algorithm first detected the error in the SK draw flow rate and then in the liquid feed flow rate. As explained earlier, this is because of the fact that the molecular weight of liquid feed is higher than that of SK. An error of 18 MT/day in liquid feed flow rate will affect less than the 13 MT/day error in SK flow in terms of lb-moles/hr and hence the error in SK flow was detected first.

If two errors were introduced in the streams like 1 and 50, 1 and 2, 11 and 12, which are connected to the same node of the process diagraph (see Figure 4.1), it would be difficult to detect such errors in both the streams. For example if the errors are in reflux flow (stream 1) and in Top CR pump around flow, the material balance will not match at nodes 1, 2 and 3. Then at node 1 the error can be due to reflux flow or due to UN flow or both. At node 2 the error can be due to Reflux flow or due to Top CR flow or both while at node 3 the error can be in Top CR flow. So it would be difficult to judge, which two streams are actually in error. Moreover, if the magnitude of error in both these streams are comparable and of opposite signs, the material balance at node 2 will hold good which will create further problems in error detection.

Three errors in flow rate measurements

The algorithm was also tested to detect more than two errors in the flow rate measurements at a time. The results obtained are shown in Table 5.3. As shown in this table, the detection of errors and the estimation of their magnitude got worse when three errors were present. The problems faced with the two errors at a time got worse when three errors were present. If any two of the streams having error were connected to the same node, it became difficult to detect the outlier.

Introduction of 4, 5 or more errors at a time was also attempted, but the results obtained were very bad. It badly affected the detection as well as the estimation of magnitude. The estimated magnitude was significantly different from the actual magnitude of error introduced.

Table 5.3 Three errors in flow rate measurements

Case No.	Flow rate perturbed simultaneously	Magnitude of error introduced (MT/day)	Results obtained: errors (in order of detection) in	Magnitude of estimated error (MT/day)
1	i) SK flow	+16	i) SK flow	+19.1
	ii) LGO CR flow	+17	ii) LGO CR flow	+14.1
	iii) LGO flow	+17	iii) LGO flow	+20.5
2	i) LGO flow	+17	i) LGO flow	+20.8
	ii) HGO CR flow	+20	ii) HGO CR flow	+16.7
	iii) HGO flow	+20	iii) HGO flow	+24.2
3	i) LGO CR flow	-18	i) LGO CR flow	-21.2
	ii) LGO flow	-17	ii) LGO flow	-14.1
	iii) HGO CR flow	-20	iii) HGO CR flow	-24.1
4	i) SCN flow	-15	i) SCN flow or SK	-22.3 or
	ii) Liquid feed flow	-21	CR flow or both	-18.2
	iii) LR flow	-22	ii) Liquid feed flow	-16.7
			iii) LR flow	-17.1

Single error in temperature measurements

One gross error was introduced at a time in the temperature measurements and GLR algorithm used for detection with enthalpy balance as the constraint as described earlier. The results are shown in Table 5.4. It was found that the algorithm was unable to detect any error in the condenser and the flash zone temperatures. Failure of algorithm to detect the error in condenser is due to the fact that the temperature of condenser in question is around 35 – 40^{0}C and in this temperature range, the correlations used for enthalpy calculation are not sensitive enough to the temperature change. Even an error of 25 to 30 ^{0}C in the condenser temperature did not change the enthalpies of the streams leaving condenser significantly and the error remain undetected.

Table 5.4 Single error in temperature measurements

Case No.	Temperature perturbed	Magnitude of the error introduced (^{0}C)		Results obtained: (error detected in)	Estimated amount of error (^{0}C)	
		+ve	-ve		+ve	-ve
1	column Top Temp	15	15	column top temp.	17	17.1
2	Top CR draw Temp	12	13	Top CR draw temp.	10.8	11.2
3	SK CR draw Temp	8	8	SK CR draw temp.	9.4	9.2
4	SK draw Temp	7	8	SK draw temp.	8.5	9.1
5	LGO CR draw Temp	8	8	LGO CR draw temp.	7.2	7.0
6	LGO draw Temp	10	10	LGO draw temp.	12.3	12.0
7	HGO CR draw Temp	7	7	HGO CR draw temp.	6.2	6.0
8	HGO draw Temp	11	12	HGO draw temp.	10.1	11.2
9	column bottom Temp	9	11	column bottom temp.	11.1	13.2

The flash zone (node 10), the change in temperature would affect the enthalpies of the streams 42 and 44. Stream 42 is the liquid leaving flash zone (plate 44) and the stream 44 is the liquid feed entering in the flash zone. If we increase the temperature by any value, it will increase the enthalpies of both streams 42 and 44. The increase in the enthalpy of stream 42 will result in increase in 'total enthalpy out' from plate 44 whereas the increase in the enthalpy of stream 44 will result in the increase in the 'total enthalpy in' to the plate 44. So any increase or decrease in enthalpy will cancel each other and it would not be possible for us to detect the error in flash zone temperature.

Two errors in temperature measurements

As described above, barring error in condenser and flash zone temperature the present algorithm was able to perform reasonably in detecting the presence of gross error in a single temperature measurement and its estimation. Subsequently two errors at a time were introduced in the temperature measurements. Of course flow measurements for these tests were assumed to be error free. The two temperatures were selected randomly. The results obtained for a few typical cases are shown in Table 5.5

Table 5.5 Two errors in temperature measurements

Case no.	Temperature perturbed	Magnitude of error introduced (^{0}C)	Result obtained: error (in order of detection) in	Magnitude of estimated error (^{0}C)
1	i) SK draw Temp	+9	i) SK draw temp.	+11.1
	ii) SCN draw Temp	+10	ii) SCN draw temp.	+11.9
2	i) Column Top Temp	+17	i) column top temp.	+19.8
	ii) HGO draw Temp	+13	ii) HGO draw temp.	+11.4
3	i) HGO CR draw Temp	-9	i) HGO CR draw temp.	-7.4
	ii) Top CR draw Temp	- 14	ii) Top CR draw temp.	-11.3
4	i) SK draw Temp	-10	i) SK draw temp.	-12.4
	ii)LGO CR draw Temp	-10	ii) LGO CR draw	-8.4
5	i)Column bottom Temp	+12	i) column bottom temp.	+14.9
	ii)LGO CR draw Temp	-10	ii) LGO CR draw temp.	-8.3
6	i) HGO draw Temp	-13	i) HGO draw temp.	-11.2
	ii) Kero CR draw Temp	+10	ii) SK CR draw temp.	+12.0

A comparison of Table 5.5 with Table 5.4 (when only one temperature was in error) shows that when two errors are present the magnitude of detectable error has increased. Also the estimated magnitude of the error has deteriorated. This is in line with the similar observation made in case of multiple errors in flow rate measurement. The order of error detection in temperature measurement is not that of larger error in temperature being detected first but it is in order of decreasing enthalpy of the streams. In case 3 (Table 5.5) the error was larger in Top CR draw temperature but error in HGO CR draw temperature was detected first because enthalpy discrepancy caused by this stream at node 8 was larger than that caused by Top CR at node 3.

Three errors in temperature measurements

To check if the detection of three simultaneous errors in temperature measurement was at all possible by using this algorithm, a few tests were made and their results are presented in Table 5.6. These three temperatures to be perturbed were chosen at random. Again the assumption that no flow rate measurement has any error is valid here.

Table 5.6 Three errors in temperature measurements

Case No.	Temperature Perturbed	Magnitude of error Introduced (^{0}C)	Results Obtained: (errors in order of detection) in	Magnitude of estimated error (^{0}C)
1	i) SK draw Temp	+11	i) SK draw temp.	+14.3
	ii)SK CR draw Temp	+11	ii) Kero CR draw temp.	+13.9
	iii) Top CR draw Temp	+15	iii) Top CR draw temp.	+12.7
2	i) SK draw Temp	-11	i) SK draw temp.	-13.8
	ii) HGO draw Temp	-16	ii) HGO draw temp.	-13.3
	iii) LGO draw Temp	-14	iii) LGO draw temp.	-17.9
3	i) Column Top Temp	+20	i) column Top temp.	+23.1
	ii)Column Bottom Temp	+14	ii)column Bottom temp.	+17.9
	iii) LGO CR draw Temp	-12	iii) LGO CR draw temp.	-10.6
4	i) HGO CR draw Temp	+10	i) HGO CR draw temp.	+7.9
	ii) LGO CR draw Temp	-12	ii) LGO CR draw temp.	-9.6
	iii) Top CR draw Temp	-15	iii) Top CR draw temp.	-12.7

As seen from the table, the results further deteriorated making the GLR method of error detection and compensation less satisfactory. Attempts were made to detect 4, 5 or even 6 errors at a time. But this resulted in very high detectable value of error and also the estimated magnitude of error were far apart from the actual error.

Simultaneous errors in flow rate and temperature measurements

In a real plant the error may be in any of the measurements, whether it is flow rate or temperature. Such situations were also created by perturbing both flow rate and temperature measurements and tested with the algorithm. In all such cases material balance errors must be eliminated first before proceeding for enthalpy balance errors. This means that all flow rate errors are detected and compensated before looking for errors in temperature measurements.

Table 5.7 shows the results for one error each in flow and temperature measurements. As before (Table 5.1), error in SCN flow can be mistaken for error in SK SR flow and the estimation in this case is very poor (case 1, Table 5.7). The magnitude of estimated for temperature in almost all the cases is less satisfactory as compared to the case when there was no error in flow measurement.

Table 5.7 One error in flow rate and one error in temperature measurements

Case No.	Flow rate and temperature perturbed	Magnitude of error introduced	Result obtained: error (in order of detection) in	Magnitude of estimated error
1	i) SCN flow ii)LGO CR draw Temp	+11MT/Day +11 ^{0}C	i) SK CR flow or SCN flow or both ii) LGO CR draw temp.	+19.7 or +12.3MT/d +9.3 ^{0}C
2	i) LGO flow ii)HGO CR draw Temp	-12 MT/Day -9 ^{0}C	i) LGO flow ii) HGO CR draw temp	-14.1 MT/d -7.6 ^{0}C
3	i) SK flow ii) SK draw Temp	+14MT/Day -10 ^{0}C	i) SK flow ii) SK draw temp.	+19 MT/d -14 ^{0}C
4	i) LR flow ii) HGO draw Temp	-18 MT/Day +11 ^{0}C	i) LR flow ii) HGO draw temp.	-20.3 MT/d +13.6 ^{0}C

Also if the flow rate and temperature belong to the same node like in Case 3 (SK draw flow rate and SK draw temperature belong to node 5), the prediction of the error is badly affected.

Tables 5.8 and 5.9 show some results when more than one error are present either in flow rate or in temperature measurements. The results are in tune with the earlier observations.

Table 5.8 Two errors in flow rate and one error in temperature measurements

Case No.	Flow Rates and Temperature perturbed	Magnitude of error	Result obtained: error (in order of detection) in	Magnitude of error estimated
1	i) SK flow ii) HGO CR flow iii) HGO draw Temp	+14MT/Day +14MT/Day +12 ^{0}C	i) SK flow ii) HGO CR flow iii) HGO draw temp.	+15.8MT/d +16.2MT/d +10.6 ^{0}C
2	i) LGO flow ii) Liquid feed flow iii)Top CR draw Temp	-13 MT/Day -19 MT/Day -15 ^{0}C	i) LGO flow ii) Liquid feed flow iii)TopCR draw temp	-15 MT/d -21.1 MT/d -13 ^{0}C
3	i) SCN flow ii) LR flow iii)column Bottom Temp	-12 MT/Day +19MT/Day +12 ^{0}C	i) Kero CR flow or in SCN draw or in both ii) LR flow iii)column bottom temp.	-21.3 or -13.5 MT/d +22.3MT/d +16.2 ^{0}C

Table 5.9 One error in flow rate and two errors in temperature measurements

Case No.	Flow rates and temperature perturbed	Magnitude of error	Result obtained: (error in order of detection) in	Magnitude of estimated error
1	i) SK flow ii)LGO CR draw Temp iii) LGO draw Temp	-12 MT/day -10 ^{0}C -12 ^{0}C	i) SK flow ii)LGO CR draw temp iii) LGO draw flow	-13.1 MT/d -7.9 ^{0}C -15.1 ^{0}C
2	i) LGO flow ii) LGO draw Temp iii)HGOCR draw Temp	+11MT/day +12 ^{0}C +9 ^{0}C	i) LGO flow ii) LGO draw temp. iii)HGO CR draw temp.	+16 MT/d +16.3 ^{0}C +7.6 ^{0}C
3	i) liquid feed flow ii) column Top Temp iii) SCN draw Temp	+17MT/day +18 ^{0}C -11 ^{0}C	i) liquid feed flow ii) column Top temp. iii) SCN draw temp.	+18.3MT/d +21.3 ^{0}C -13.2 ^{0}C

Chapter 6

CONCLUSIONS AND RECCOMANDATIONS

In this work an attempt have been made to solve steady state gross error detection problem for a crude distillation unit. In the absence of redundancy of measurements, it has not been possible to eliminate random errors using a data reconciliation algorithm. Generalized Likelihood Ratio method has been used for gross error detection. In any oil refinery only a few stream flow rates and temperatures are measured and no measurements are made on the composition at any location. Some reasonable assumptions were made to calculate the composition and other parameters at different locations of the column using a simulator. This enabled the detection of gross errors in flow rates as well as in temperatures. The algorithm was tested using several sets of real plant data offline. In some of the measurements errors were deliberately introduced. Gross errors in maximum of three temperatures or three flow rates or both were detected successfully in almost all the cases investigated. The minimum amount of error that can be detected increases as the number of errors increase.

Several assumptions had to be made in the development of the present algorithm because of very limited measured data being available. It is, therefore, necessary that the algorithm should be tested extensively using real time plant data. It has been concluded by many researchers that any single method can not guarantee detection of all the errors successfully. This is so because of the fact that almost all gross error detection algorithms are statistical in nature and there is a finite probability that a given error escapes undetected by any one method. It is, therefore, important that one or more other methods of gross error detection such as Principal Component Analysis or the methods based on Artificial Neural Networks should also be worked out and coupled with GLR method in order to make the package more robust and reliable.

REFERENCES

Almasy G.A. and R.S.H. Mah, Estimation of Error Variances from Process Data, *Ind. Eng. Chem. Process Dev.*, **23**, 779-784 (1984).

Biegler L.T. and Tjoa B., Simultaneous Strategies for Data Reconciliation and Gross Error Detection of Nonlinear Systems, *Comput. Chem. Engg.*, **15**, 679-690 (1991).

Crowe C. M., Campos Y. A. G. and Hrymak A., Reconciliation of Process Flow Rates by Matrix Projection, Part I: Linear case, *AIChE J.*, **29**, 881-888 (1983).

Crowe C. M., The Maximum-Power Test for Gross Errors in the Original Constraints in Data Reconciliation, *The Can. J. Chem. Eng.*, **70**, 1030-1036 (1992)

Crowe C.M. and Tong H., Detection of Gross Errors in Data Reconciliation by Principal Component Analysis, *AIChE J.*, 41, 1712-1722, (1995)

Deo N., *Graph Theory with applications to Engineering and Computer Science*, Prentice Hall, Englewood Cliffs, NJ (1974).

Dutta P., *Steady state Data Reconciliation and Gross Error Detection For a Crude Distillation Unit*, M. Tech. Dissertation, IIT Kanpur, Dept. of Chem. Engg., March 2000

Edgar T.F. and Himmaelblau D.M., *Optimization of Chemical Processes*, Mc.Graw-Hill Book Company (1988).

Edmister W. C. and Lee B.I., *Applied Hydrocarbon Thermodynamics*, Vol 1, Gulf publishing company, Houston, Texas (1984)

Himmelblau D.M., *Process Analysis by Statistical Methods*, John Wiley, Austin, Texas (1970).

Jao L., Ten R. and Yang Y., A Study of Gross Error Detection and Data Reconciliation in Process Industries, *Comput. Chem. Engg.*, **19**, S217-S222 (1995)

Jiang Q. and Bagajewicz M.J., Gross Error Modelling and Detection in Plant Linear Dynamic Reconciliation, *Comput. Chem. Eng.*, 22, 1789-1801 (1998)

Kesler M.G., Lee B.I., Improve Prediction of Enthalpy of Fractions, *Hydrocarbon Process*, **55(3)**, 153-163 (1976)

Mah R.S.H., Stanley G.M. and Downing D.M., Reconciliation and Rectification of Process Flow and Inventory Data, *Ind. Engng. Chem. Process Des. Dev.*, **15**, 175-183 (1976)

Mah R.S.H., *Chemical Process Structures and Information Flows*, Butterworth, Stoneham, (1990).

Mah R.S.H. and Tamhane A.C. Detection of Gross Errors in Process Data, *AIChE J.*, **28**, 828-839 (1982).

Mah R.S.H and Tamhane A.C., Data Reconciliation and Gross Error Detection in Chemical Process Networks, *Technometrics*, **27**, 409-422 (1985).

Narasimhan S. and Mah R.S.H, Generalized Likelihood Ratio Method for Gross Error Detection, *AIChE J.*, **33**, 1514-1521 (1987).

Prasanth K.P., *Modelling of Crude Distillation Unit for Online Applications*, M. Tech. Dissertation, IIT Kanpur, Dept. of Chem. Engg., September 2000

Ramagnoli J., Tonelli S. and Schbib S., Gross Measurements Error Detection/Identification For An Industrial Ethylene reactor, *Comput. Chem. Engg.*, **20**, S1559-S1564 (1996)

Serth R.W. and Heenan W.A., Gross Error Detection and Data Reconciliation in Steam-Metering Systems, *AIChE J.*, **32**, 733-742 (1986).

Willsky A.S. and H.L. Jones, A generalized Likelihood Ratio Approach to State Estimation in Linear Systems Subject to Abrupt Changes, *Proc. IEEE Conf. Decision and Control*, 846-856 (1974).

Appendix A

Data used for Q matrix estimation for example problem

The five columns represent the five variables in the problem and the fifty rows are the randomly perturbed pseudo-measurements.

(1)	(2)	(3)	(4)	(5)
226.6520	324.3390	323.7090	224.4760	100.4120
224.2460	325.7180	323.8380	224.9700	100.5030
225.6010	326.0580	326.4750	224.6800	97.6980
224.9600	325.0490	324.2650	224.3830	98.8720
226.4960	325.8180	325.9340	225.4080	100.5040
225.4290	325.6540	324.4520	227.4280	99.2920
225.6690	325.4810	326.1270	225.3330	99.6640
225.8820	325.4180	326.9880	224.3100	99.1870
226.6370	324.5680	325.8360	225.7900	100.2660
225.6560	324.8290	325.9730	225.8390	100.5510
223.9990	324.1540	325.0100	225.4130	99.9770
224.9800	323.3570	323.7930	223.5020	99.3700
224.3030	324.6990	326.1470	226.2390	101.9800
225.3720	325.3900	324.1960	225.5190	99.5580
225.3190	324.9380	323.9030	226.2290	100.4290
225.2810	325.2560	327.3970	222.9300	99.1700
225.3900	324.9530	325.5990	224.2520	100.0030
224.8610	323.4150	324.7930	225.9610	100.7790
223.5510	324.4920	324.9440	223.9620	98.4220
224.9490	323.8100	325.1450	224.5560	99.2520
224.9760	325.4810	325.5710	225.1500	99.3530
223.4910	323.8890	324.7790	225.9520	100.6910
224.0480	324.7910	327.1760	224.0860	99.6600
224.3850	326.4160	324.5490	225.3930	99.4250
223.8770	325.5300	325.6040	224.6220	99.3620
225.0140	325.1080	324.0570	226.7340	101.6300
226.7590	325.8880	324.1760	226.2570	99.6040
224.1960	324.2520	323.4170	224.3420	100.5760
224.4570	326.6920	325.9470	224.2670	100.9250
225.2000	325.0250	322.9330	223.7860	100.8100
225.3810	324.0920	326.3930	225.3850	99.4790
223.4210	327.0900	325.7030	226.8140	100.4110
224.3590	325.2770	324.3250	226.0030	97.3370
225.1770	324.9220	324.7710	225.9120	101.1240
223.4220	323.5900	324.8720	225.0970	99.9600
224.5780	325.0280	325.6490	224.9770	100.2300
225.0110	324.0690	325.0670	224.7410	100.4800
226.0980	324.6310	324.7890	226.8480	100.2150
223.9430	326.8670	325.0090	225.4570	99.0980
225.2100	327.2710	325.3220	226.4840	100.3630
225.1340	324.2960	324.4280	225.2990	98.9300
226.3650	323.8730	326.5110	224.5790	99.3090

225.2030	324.6350	324.5010	225.6210	97.6170
225.0230	323.8770	324.4230	225.9450	98.6900
225.0760	323.7460	324.1250	225.2370	100.0840
224.3420	324.1240	326.0760	224.1130	99.0800
224.2210	325.3920	326.5510	225.3600	99.3060
223.4180	323.9640	324.1420	226.2540	100.4100
225.9860	324.0210	325.5780	226.1470	99.0160
223.7460	323.0660	324.6080	224.6510	100.7810

Appendix B

A Brief Review of Graph Theory

In this section we present some key concepts of graph theory, which have been extensively used in our work. This description is essentially drawn from Deo (1974) and Mah (1990).

Graph and Subgraph

A undirected (or directed) graph $G(V,E)$ consists of a set of objects $V = \{v_1, v_2, ..., v_n\}$ called vertices or nodes and another set $E\{e_1, e_2, ..., e_m\}$called edges, such that each edge e_k is identified with an unordered (or ordered) pair (v_i, v_j) which are called end nodes. The edges are said to be incident on these nodes. Note that by its very definition, a graph contains both the end nodes of every edge it contains. Schematically, nodes are represented as circles and arcs joining these circles represent edges.

Thus, the flowsheet of a process plant can be represented as a graph with the nodes corresponding to the process units and the streams corresponding to the edges. In addition a hypothetical node is used to represent the environment on which the feed and the products streams are incident. An illustration of the process network and its corresponding graph are given in Figures A.1 and A.2 respectively. In Figure A.2 node 6 is the environmental node. For process network all graphs are directed graphs. A directed graph is also called diagraph.

A graph $G'(V', E')$ is said to be a subgraph of $G(V, E)$ if $V' \in V$ and $E' \in E$ and each edge of G' has the same end vertices as in G. For example, the graph shown in Figure A.3 is a subgraph of $G(V, E)$ shown in Figure A.2.

Path Cycles and Connectivity

A path between vertices v_0 and v_l is an alternating sequence of distinct vertices and edges $v_0 \ e_0 \ v_1 \ e_1 ..... \ v_{l-1} \ e_{l-1} \ v_l$ where (v_i, v_{i+1} are the end nodes of edge e_i). If $v_0 = v_l$ then the path is called a cycle. For example, in Figure A.2, the sequence of edges *1*, *2* and *3* together with their end nodes is a path and the sequence *1*, *2*, *3*, *5* and *6* together with their end nodes is a cycle.

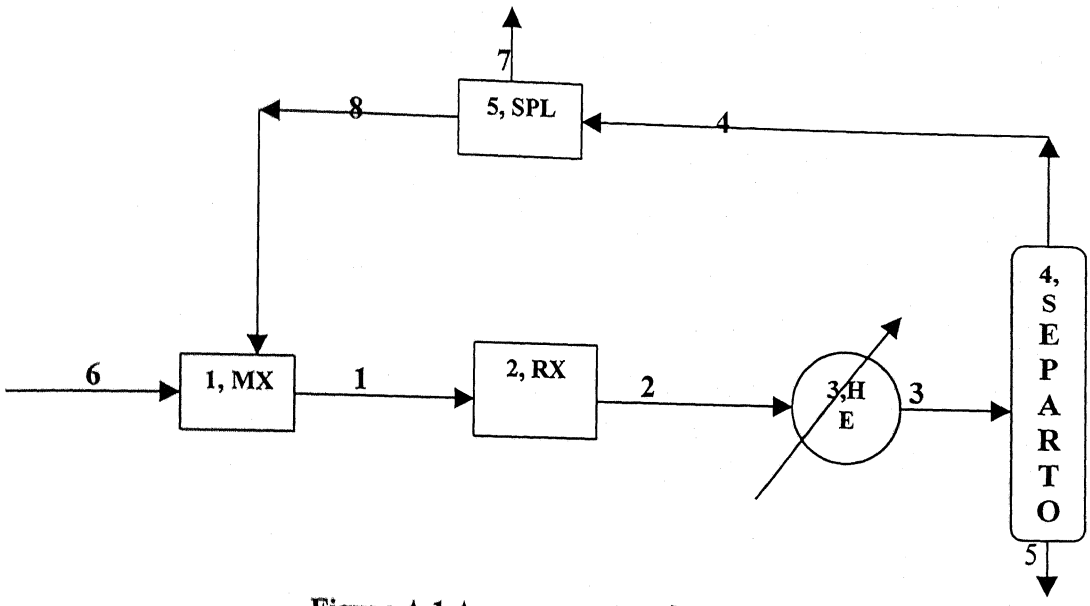


Figure A.1 A process network

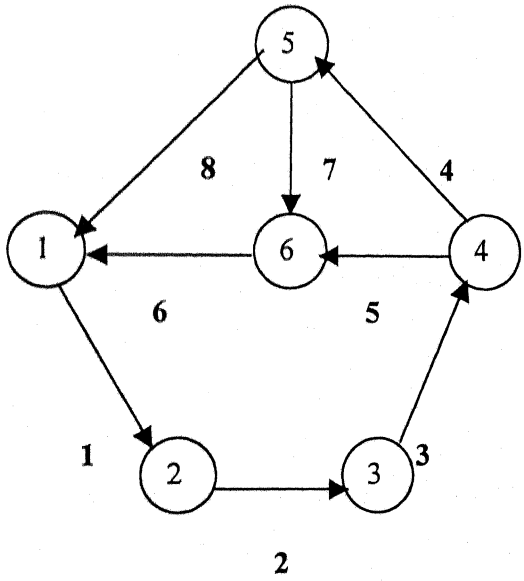


Figure A.2 Graph G

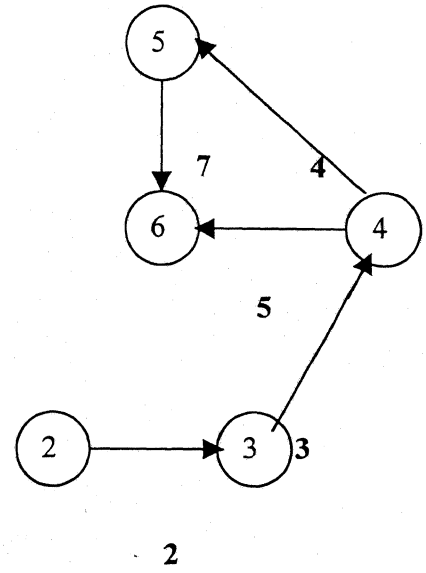


Figure A.3 Subgraph of G

A graph is said to be connected if there is at least one path between every pair of vertices in G. The graph in Figure A.2 is connected. However, the graph in Figure A.4 is disconnected. For a process network all graphs are connected together. Each connected

portion of a disconnected graph is known as component. The disconnected graph shown in Figure A.4 has two components.

Trees, Spanning trees, Branches and Chords

A tree is a connected graph that does not contain any cycle. The graph shown in Figure A.5 is a tree. A tree T, is said to be a spanning tree of a graph G, if it is a subgraph of G and all vertices of G are also contained in T. For example, the graph shown in Figure A.6 is a spanning tree of the graph in Figure A.2, whereas that shown in figure A.5 is not. An edge in a spanning tree T is called a branch of T, while an edge of G which is not in T is called chord. Note that the branches and chords are defined with respect to a spanning tree. For example, edges 1, 3, 4, 7, and 8 shown in the spanning tree of Figure A.6 are branches while edges 2, 5 and 6 which are present in Figure A.2 but not in spanning tree of Figure A.6 are chords.

Cutsets and fundamental cutsets

A cutset of a connected graph G, is a set of edges whose removal from G disconnects it, but the removal of proper subset of these edges does not disconnect G. For example in Figure A.2, the set of edges 3, 6, 8 is a cutset. However, edges 2, 3, 6, 8 does not form a cutset (although, their removal disconnects G) since the removal of a proper subset of edges 3, 6, and 8 itself can disconnect G.

Fundamental cutsets are defined with respect to a spanning tree T of graph G. A fundamental cutset is a cutset of graph G which contains exactly one branch of tree T. For example, in Figure A.2, edges 2, 6 and 8 form a fundamental cutset with respect to the spanning tree in figure A.6 where edge 8 is a branch and all the remaining edges are chords. On the other hand the set of edges 4, 5, 6 and 8 is not a fundamental cutset with respect to the spanning tree shown in Figure A.6 (although, it is a cutset) since it contains more than one branch of T.

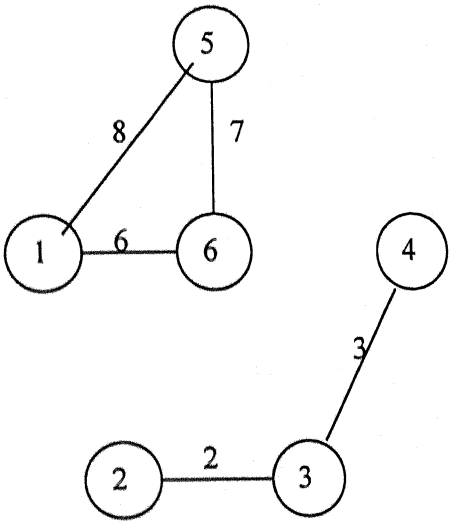


Figure A.4 A disconnected graph with two components

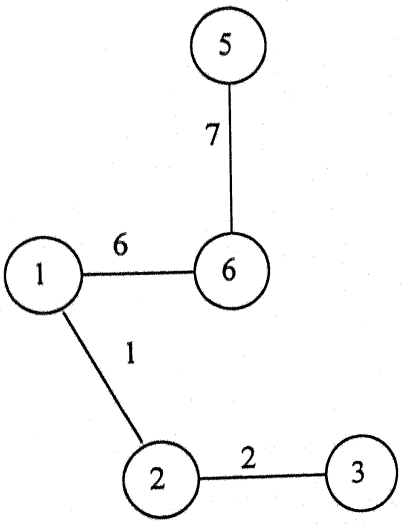


Figure A.5 A tree

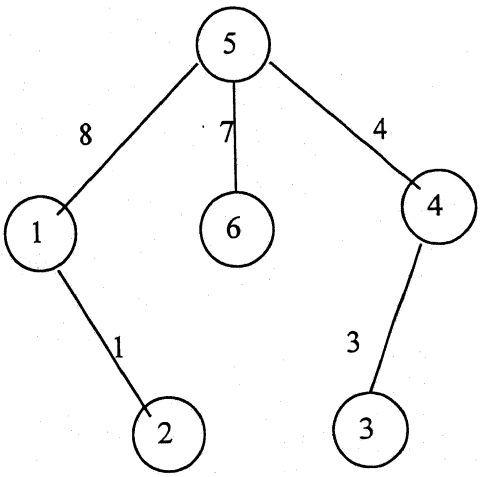


Figure A.6 A spanning tree of G

Matrix Representation of Digraphs and Graphs

Matrices are used to represent a process digraph. One obvious matrix representation is to denote each vertex by a row and each edge by a column. The (i,j)th element of this matrix will be assigned the value "+1", if edge j is incident into vertex i, "-1 ", if edge j is incident out of vertex i, and "0", if vertex i is not connected to vertex j. Such a matrix is called *incidence matrix*, M'. Figure A.8 shows the incidence matrix for the digraph shown in Figure A.7

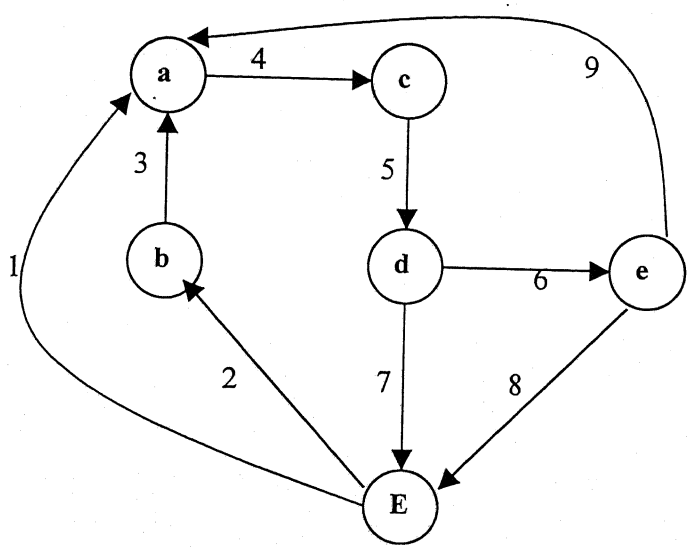


Figure A.7 A Process Digraph

$$M' = \begin{array}{c|ccccccccc} & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ a & 1 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 1 \\ b & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ c & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ d & 0 & 0 & 0 & 0 & 1 & -1 & -1 & 0 & 0 \\ e & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & -1 \\ E & -1 & -1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{array}$$

Figure A.8 Incidence matrix, M'

It can readily be shown that for a connected digraph, the rank of its incidence matrix is always N-1 (N is number of rows of the incidence matrix). We may therefore, omit one row of M', without any loss of information. The omitted vertex is known as reference vertex or the datum node, and the matrix reduced by one row is called *reduced incidence matrix*. Reduced incidence matrix occurs most commonly in material and energy balances around the network.

Observability and Redundancy

In a process flow network there are several variables, some of them are measured and some of them are not measured, for the reasons of cost, convenience or technical feasibility. Depending upon the structure of the process network, unmeasured variables can be estimated from the measured variables without violating material and energy balance constraints. If we are able to change a value of the variable without violating constraints then the change is said to be feasible (data reconciliation). This means that values of some other related variables would have to be changed appropriately so the constraints remain satisfied.

For a given process network and given set of instrument placements, if we can make a feasible change for a variable without being detected (or observed) by the instruments, then the variable is said to be unobservable. By definition, a measured variable is certainly observable, but an unmeasured variable may or may not be observable. For instance, in a tank with one inflow and one outflow, we may apply feedback control to the liquid level by manipulating one or two flow rates if the level is measured. If we measure the flow rates but omit the level measurement, the level is still observable but control system is now forward feed rather than back feed.

If we delete a measurement associated with a given variable, and if the variable remains observable, then the measurement is said to be redundant. In a process network, we require certain variables to be observable, others to be measured but not redundant, and still other measurements to be redundant. For data reconciliation we need much redundancy in the measurements.

Appendix C

Users Manual

The whole program of gross error detection is divided into three parts

1. Q matrix generation
2. Gross error detection in flow rates
3. Gross error detection in temperatures

Q matrix generation

All the programs for Q matrix generation are listed in the directory 'Qmatrix'. There are four subroutines in this program

1. true.f90: It is the main program which calculates the average (true) value of the variables over the given number of data sets and then generates fifty random data sets based on the true value.
2. qmat.f90: This subroutine calculates the Q matrix for the given data sets.
3. hmat.f90: This subroutine calculates the covariance matrix of residuals ($H = AQA^T$).
4. inv.f90: This subroutine calculates the inverse of H matrix. This H^{-1} is required in the GLR method.

Input files

1. input.dat: This file contains a number of consecutive data sets of the measured variables, which are averaged to obtain 'true' values.
2. a.dat: This file contains the incidence matrix of the process, A.

Output files

1. true_value.out: This file contains the average (true) values of the measured variables.
2. random_variate.out: This file contains the fifty data sets used for the estimation of Q matrix.
3. Qmatrix.out: This file contains the Q matrix.
4. H_mat.out: This file contains the H matrix.
5. ainv.out: This file contains the inverse matrix of H.

Procedure

1. Enter all the data sets in the file input.dat.
2. Run the program to get the true values, Q matrix and inverse of H matrix.

Gross error detection in flow rates

All the programs for gross error detection in flow rates are listed in directory 'new1'. There are five subroutines in this program.

1. mass.for: This is the main program. It reads all the flow rates (measured and calculated) and converts the measured flow rates into lb-moles/hr.
2. glr.for: This subroutine performs the material balance at all the nodes and then performs the GLR test to detect the gross error.
3. number.for: This subroutine tells us which measured stream is in error.
4. magni.for: This subroutine calculates the magnitude of error in lb-moles/hr.
5. mag1.for: This subroutine converts the magnitude of error in MT/day.

Input files (supplied by user)

1. a.dat: This file contains the incidence matrix of the process.
2. ainv.dat: This file contains the inverse of the H matrix (from Q matrix estimation).
3. evector.dat: This file contains fifty unity vectors (length of each vector is fifty). GLR method uses these vectors to detect the error.
4. input.dat: This file contains the twelve measured flow rates in MT/day.

Input files (calculated)

1. mol_wt.dat: This file contains the molecular weight of the measured streams. These molecular weights are obtained by running the CDU simulator.
2. glr-dat1.dat: This file contains the flow rates of the unmeasured streams. These flow rates are calculated with the help of CDU simulator.

Output files

1. enthy1.out: This file contains flow rates of all the fifty streams in the order according to the process diagraph.
2. results.out: This file contains the results of the GLR test.

Procedure

1. Run the CDU simulator for a number of consecutive sets of data and each time save the file glr-dat1.out (flow rates of all the intermediate liquid and vapor streams) by a different name. Arrange these measured and unmeasured flow rates in the order as described in the process diagraph. Now these data sets will be used for calculating the 'true' data set and to calculate the Q matrix and the inverse of H matrix (directory 'Qmatrix').

2. Put the average value of unmeasured flow rates in the file glr-dat1.dat ('New2').
3. Get the molecular weight of all the measured streams and put these in file mol-wt.dat.
4. Now enter all the values of measured flow rates in the file input.dat and run the program mass.for to detect the gross errors in measurements.
5. If the operating conditions changes significantly, we have to again repeat the steps 1 to 3.

Gross error detection in temperatures

All the programs for gross error detection in temperatures are listed in the directory 'New4'. There are seven subroutines in this program

1. temp.for: This is the main program which reads all the measured and unmeasured flow rates and temperatures and calculates the enthalpy of all the streams involved.
2. enth1.for: This subroutine contains the enthalpy correlation for the liquid streams.
3. enth2.for: This subroutine contains the enthalpy correlations of the vapour streams.
4. glr.for: This subroutine performs the enthalpy balance around each node and then performs the GLR test to detect the error.
5. plate.for: This subroutine tells that which temperature is in error.
6. magni.for: This subroutine calculates the amount of error in terms of enthalpy discrepancy.
7. new.for: This subroutine calculates the amount of error in terms of temperature.

Input files (supplied by user)

1. a.dat: This file contains the incidence matrix of the process.
2. ainv.dat: This file contains the inverse of H matrix.
3. evector.dat: This file contains fifty unity vectors (length of each vector is fifty). GLR method uses these vectors to detect the error.
4. temp.dat: This file contains the values of measured temperatures.
5. flow.dat: This file contains the values of measured flow rates.

Input files (calculated)

1. liq_comp.dat: This file contains the composition of some liquid streams ('Avg5')

2. vap_comp.dat: This file contains the composition of some vapor streams ('Avg6')
3. property.dat: This file contains the properties (molecular weight, specific gravity, acentric factor, critical temperature) of the pseudo-components (cdu-flas.out)
4. sp_liq.dat: This file contains the specific enthalpies of some liquid streams ('Avg4')
5. sp_vap.dat: This file contains the specific enthalpy of some of vapor streams ('Avg4')
6. mol_wt.dat: This file contains the values of molecular weights of measured flow rates.

Output files

1. enthy.out: This file contains the enthalpy of all the fifty streams in the order according to the process diagraph.
2. results.out: This file contains the results of GLR test.

Procedure

1. Run the CDU simulator for a number of consecutive sets of data and each time save the file glr-dat1.out (flow rates of all the intermediate liquid and vapor streams), glr-dat3.out (enthalpies of all the intermediate liquid and vapor streams), glr-dat4.out (enthalpies of all the side streams), glr-dat5.out (specific enthalpy of all the liquid and vapor streams), glr-dat6.out (composition of all the liquid streams) and glr-dat7.out (composition of all the vapor streams) by a different name for a different data set. For example, for flow rates of intermediate streams calculated using glr-dat1.out are saved in glr-dat1_*i*.out where $i = 1,\ldots,6$ if six different sets are used in the averaging. The average of these are found in directory 'Avg1'. Similarly other averages are found using directory 'Avg2', 'Avg3', 'Avg4', 'Avg5' and 'Avg6' respectively. Now from the directories 'Avg2' and 'Avg3', we can get the enthalpies of all the fifty streams in question. Arrange them in the order as described in the process diagraph. Now take these fifty enthalpies as the true set of data and estimate the Q matrix and the inverse of H matrix (directory 'Qmatrix').
2. Take the average values of compositions of liquid and vapor streams from directories 'Avg5' and 'Avg6' respectively. Copy these values in files liq_comp. and vap_comp.dat respectively.

3. Take the average values of specific enthalpies of liquid and vapor streams. from directory 'Avg4' and copy these values in the files sp_liq.dat and sp_vap.dat accordingly.
4. Copy the H matrix from directory 'Qmatrix' and paste it in the file ainv.dat.
5. Copy the file cdu-flas.out from the CDU simulator to the file property.dat.
6. After detecting the errors in flow rates, run the simulator for the corrected measurements.
7. Flow rates of intermediate liquid and vapor streams are directly read from the CDU simulator.
8. Enter the measured flow rates in flow.dat (in MT/day).
9. Enter the measured temperatures in temp.dat.
10. Enter the molecular weight of the measured streams in mol_wt.dat.
11. Run the program temp.for to detect the gross error in temperature.
12. If there is any major change in the operating conditions, repeat steps 1 to 5.

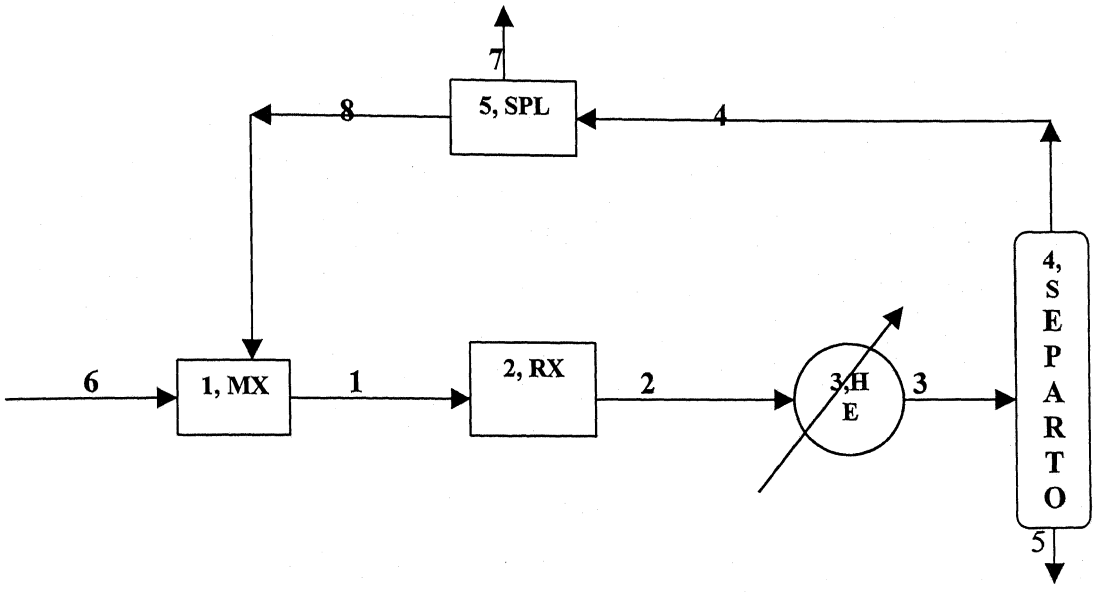


Figure A.1 A process network

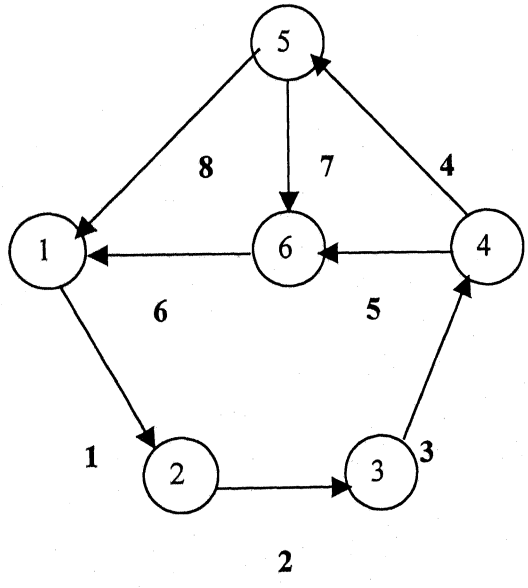


Figure A.2 Graph G

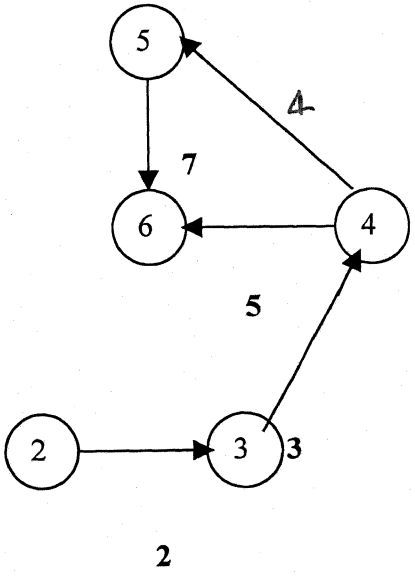


Figure A.3 Subgraph of G

A graph is said to be connected if there is at least one path between every pair of vertices in G. The graph in Figure A.2 is connected. However, the graph in Figure A.4 is disconnected. For a process network all graphs are connected together. Each connected

Matrix Representation of Digraphs and Graphs

Matrices are used to represent a process digraph. One obvious matrix representation is to denote each vertex by a row and each edge by a column. The (i,j)th element of this matrix will be assigned the value "+1", if edge j is incident into vertex i, "-1 ", if edge j is incident out of vertex i, and "0", if vertex i is not connected to vertex j. Such a matrix is called *incidence matrix*, M'. Figure A.8 shows the incidence matrix for the digraph shown in Figure A.7

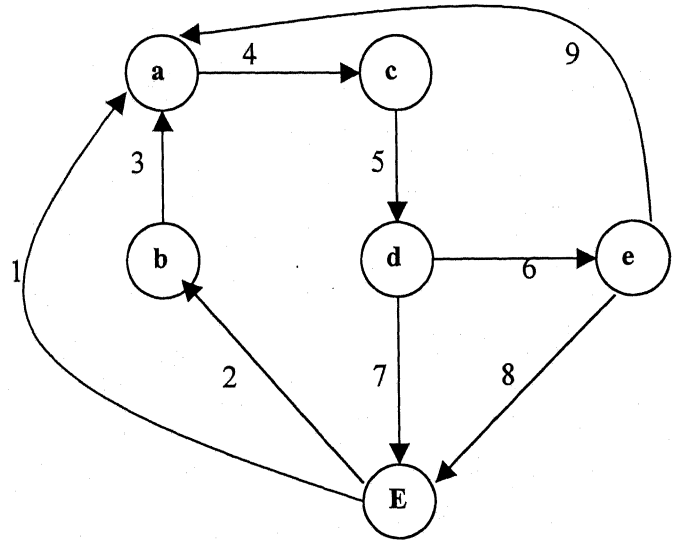


Figure A.7 A Process Digraph

$$M' = \begin{array}{c|ccccccccc} & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ a & 1 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ b & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ c & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ d & 0 & 0 & 0 & 0 & 1 & -1 & -1 & 0 & 0 \\ e & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & -1 \\ E & -1 & -1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{array}$$

Figure A.8 Incidence matrix, M'

Appendix D

Program Listing

The listing of all the computer programs used in the present work for gross error detection in steady state on-line measurements on a crude distillation unit are available with Prof. D.N. Saraf, Department of Chemical Engineering, IIT Kanpur.